Comparison of the Mitochondrial Genomes of the Common Bed Bug (*Cimex lectularius*), Eastern Bat Bug (*Cimex adjunctus*), and Swallow Bug (*Oeciacus vicarius*)

A thesis submitted to the College of

Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of Master of Science

in the Department of Biology

University of Saskatchewan

Saskatoon, Saskatchewan, Canada

By

James S. Armstrong

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

DISCLAIMER

Reference in this thesis to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Biology

University of Saskatchewan

112 Science Place

Saskatoon, Saskatchewan S7N 5E2

Canada

ABSTRACT

Species within the family Cimicidae (bed bugs) are hematophagous ectoparasites of mammals and birds. Many cimicids are of socio-economic importance. Despite the global resurgence of these pests, there is currently a paucity of information regarding the mitochondrial (mt) DNA sequences of cimicids. Therefore, I used a PCR-based primer walking strategy to amplify and sequence the near complete mitogenome of the common bed bug (*Cimex lectularius*), and several mitochondrial gene regions of the Eastern bat bug (*Cimex adjunctus*) and swallow bug (*Oeciacus vicarius*). I compared the mitochondrial genetic variability between *C. lectularius* from two populations to look for molecular markers useful for population genetic studies. Furthermore, the mt DNA sequences of these species of medical and veterinary importance were compared to those of other heteropterans to infer the evolutionary relationships of species in the family Cimicidae.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Neil Chilton for his guidance and support through both my B.Sc and M.Sc research. His expertise and mentorship during the past four years has been greatly appreciated. I would also like to thank Dr. Art Davis, Dr. Jose Andres, Dr. Dwayne Hegedus for acting as my advisory committee and external examiner. Within the department I would like to acknowledge Dr. Carlos Carvalho and Kunal Baxi for assisting me with photographs. Lastly, I would like to thank my lab mates, particularly Dr. Clare Anstead, who oversaw my training in lab techniques.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

BLAST	Basic Local Alignment Search Tool
BPS	Base Pairs
CI	Consistency Index
COI	Cytochrome Oxidase 1
COII	Cytochrome Oxidase 2
COIII	Cytochrome Oxidase 3
CR	Control Region
CYTB	Cytochrome Oxidase B
DHU	Dihydrouridine
IUB	International Union of Biochemistry
IUPAC	International Union of Pure and Applied Chemistry
lrRNA	Large Ribosomal Subunit RNA
MP	Maximum Parsimony
mtDNA	Mitochondrial DNA

ND1	Nicotinamide Adenine Dinucleotide dehydrogenase subunit 1
ND2	Nicotinamide Adenine Dinucleotide dehydrogenase subunit 2
ND3	Nicotinamide Adenine Dinucleotide dehydrogenase subunit 3
ND4	Nicotinamide Adenine Dinucleotide dehydrogenase subunit 4
ND4L	Nicotinamide Adenine Dinucleotide dehydrogenase subunit 4L
ND5	Nicotinamide Adenine Dinucleotide dehydrogenase subunit 5
ND6	Nicotinamide Adenine Dinucleotide dehydrogenase subunit 6
NJ	Neighbor Joining
Numt	Nuclear mitochondrial DNA
ORF	Open Reading Frame
PCG	Protein Coding Gene
PCR	Polymerase Chain Reaction
RI	Retention Index
RCF	Relative Centrifugal Force
rDNA	Ribosomal DNA

srRNA	Small Ribosomal Subunit RNA
Tm	Primer melting temperature
tRNA	Transfer RNA

1. INTRODUCTION

1.1. Biology of *Cimex lectularius* and other cimicids

1.1.1 Overview

The cimicids (Family Cimicidae) are blood-feeding ectoparasites of bats, birds and humans (Usinger 1966). The family Cimicidae is within the order Hemiptera and suborder Heteroptera, which are usually referred to as "true bugs" (Usinger, 1966). Most heteropterans are phytophagous and use their sucking-piercing mouthparts (rostrum) to imbibe phloem sap; however, some species have become adapted to predaceous or parasitic feeding (Weirauch et al. 2011). "Bed bugs" is a term often applied to the approximately 90 species within the Cimicidae; of these, only *Cimex lectularius*, the common bed bug, and *C. hemipterus*, the tropical bed bug, show a strong host preference for humans (Usinger 1966). Hereafter, "bed bug" will be used to refer to only *C. lectularius*. *Cimex hemipterus* can only be found in tropical or sub-tropical latitudes, while *C. lectularius* can be found anywhere humans reside (Usinger 1966; Dogget et al. 2012). Although these two species appear very similar morphologically, they can be differentiated by differences in the shape and size of the pronotum (Usinger 1966).

Cimicids have flat, shaped bodies roughly 5-10 mm in length, and reduced wings (Usinger 1966). Their eggs are approximately 5 mm and are oblong-oval with a slanting anterior cap, resembling grains of rice (Usinger 1966). Cimicids use their sucking-piercing mouthparts to penetrate their host's skin and imbibe their blood meal (Usinger 1966). The more than 90 species of cimicids are placed within 22 genera and 6 subfamilies. The most common and cosmopolitan subfamily is the Cimicinae, which includes the genera *Cimex* and *Oeciacus*. Members of the Cimicinae are morphologically delineated from

species in other subfamilies of the Cimicidae by the presence of minute serrations on pronotal bristles, and a metasternum forming a flat plate between the coxae (Usinger 1966).

It has been proposed that *C. lectularius* first evolved as a result of host switching from bats or birds to humans at a time when all three shared habitation in caves (Usinger 1966; Reinhardt and Siva-Jothy 2007; Booth et al. 2015; Zorilla-Vaca et al. 2015). There is paleoecological evidence based on examination of fossils in Egypt, suggesting that bed bugs have been parasitizing humans for at least 3,550 years (Panagiotakopulu and Buckland 1999; Delaunay et al. 2010).

1.1.2 Finding and feeding from hosts

Obligate hematophagy is one of the core adaptations that distinguish cimicids from their relatives, suggesting it likely only evolved once (Reinhardt and Siva-Jothy 2007). It is not known whether this adaptation evolved from phloem sap feeding or from feeding on fur and feathers (Reinhardt and Siva-Jothy 2007). A blood meal is required by every life stage, including adults of both sexes. Nymphs must feed within days of hatching or die, and each instar must feed before moulting. Moreover, females require a blood meal before producing eggs, and *C. lectularius* males show sexual preference towards recently fed females (Reinhardt and Siva-Jothy 2007).

Cimicids have a narrow host range relative to other insect hematophages (Reinhardt and Siva-Jothy 2007). The most common hosts are bats, swifts and swallows (Reinhardt and Siva-Jothy 2007). Hence, all hosts are warm blooded and live in social groups that have relatively immobile shelters, such as caves, buildings, and nests (Reinhardt and Siva-Jothy 2007). These hosts have also historically had overlapping habitation, a feature that lends itself to host switching by the hematophage (Reinhardt and Siva-Jothy 2007).

Cimicids locate their host by sensing CO_2, heat, and kairomones, though these stimuli have only been demonstrated to act as lures at distances less than a few meters (Wang et al. 2009). They must crawl to their host, as they lack wings and jumping ability (Usinger 1966). Unlike many other ectoparasites, cimicids do not live on their host and often only have contact with their host while feeding (Usinger 1966).

Cimicids, like other heteropterans, have mouthparts that are modified for piercing and sucking, their maxillary and mandibular stylets fused together to make the two separate canals; the feeding and the salivary (Usinger 1966). While sucking blood through their food canal they inject proteolytic enzymes, anticoagulants, vasodilatory compounds such as nitrophorin, and an unidentified anaesthetic through their salivary canal (Usinger 1966; Dogget et al. 2012; Williams and Willis 2012). When feeding on humans, bed bugs tend to feed from the neck, face, hands, and arms and other parts of the host that are exposed while sleeping (Delaunay et al. 2011; Williams and Willis, 2012).

Under *ad libitum* conditions, *C. lectularius* adults feed approximately once every few days (Usinger 1966; Reinhardt and Siva-Jothy 2007; Dogget et al. 2012). Though not strictly nocturnal, *C. lectularius* generally feeds while their host is least active. It takes roughly 10 minutes for a bed bug to fully engorge once feeding commences (Dogget et al. 2012).

1.1.3 Reproduction and life cycle

Bed bugs usually live gregariously in dark hidden spaces called harborages (Reinhardt and Siva-Jothy 2007). Box springs, mattress seams and curtains are common harborages (Dogget et al. 2012). These locations provide shelter, ample opportunity for mating, and mitigate desiccation (Reinhardt and Siva-Jothy 2007). These harborages are comprised of all life stages, and maintained by aggregation pheromones (Reinhardt and Siva-Jothy 2007; Gries et

al. 2015). Bed bugs also produce alarm pheromones when their harborages are disturbed, which is reported to smell unpleasantly sweet or like rancid strawberries (Dogget et al. 2012). *Cimex lectularius* has been observed to survive periods of starvation for months, and in temperatures near 10 ºC with favourable humidity they can live over a year (Usinger 1966).

Another adaptation distinctive of cimicids is traumatic insemination and the female paragenital system (Usinger 1966). Rather than use the female's genital opening, male cimicids use their copulatory (intromittent) organ to puncture the female's body wall and release their sperm into the abdomen (Usinger 1966; Stutt and Siva-Jothy 2001). It is thought that this behavior is a result of sexually antagonistic coevolution, due to female rejection of potential mates (Stutt and Siva-Jothy 2001; Reinhardt and Siva-Jothy 2007). The female paragenital system is thought to have evolved in response to traumatic insemination, as abdominal piercings have a deleterious effect on female survival (Usinger 1966; Stutt and Siva-Jothy 2001; Reinhardt and Siva-Jothy 2007). The cimicid paragential system includes the spermalege, an organ unique to this group of heteropterans (Stutt and Siva-Jothy 2001; Reinhardt and Siva-Jothy 2007). It has been shown that the female directs the male's intromittent organ into the spermalege, which receives the sperm. Traumatic insemination into this organ reduces the costs of wounding and mitigates the effect of pathogens (Reinhardt and Siva-Jothy 2007). Female cimicids still maintain their ancestral reproductive tract, which they use during oviposition (Usinger 1966).

With only two exceptions, cimicids have five nymphal instars, each one being bigger than the last (Usinger, 1966). The length of time it takes for a bed bug to complete its life cycle varies greatly with environmental temperature and humidity (Usinger, 1966). At a temperature of 23 ºC and 90% humidity, *C. lectularius* females lay approximately 90-150

eggs over 18 weeks, whereas *C. hemipterus* females lay roughly 50 eggs over their life span (Dogget, et al. 2012). If regular feeding is maintained, a bed bug will mature from egg to adult in roughly 50 days at room temperature (Dogget et al. 2012).

1.2 Significance of *Cimex lectularius*

Bed bugs are of particular medical, veterinary, and economic significance (Dogget et al. 2012; Zorilla-Vaca et al. 2015). It was thought that bed bugs are not vectors of disease causing agents. However, recent research suggests they are competent vectors of *Trypanosoma cruzi*, the causative agent of Chagas disease (Salazar et al. 2014). Moreover, they carry pathogens on their body and the lesions from their bites are an entry point for infection (Lowe and Romney 2011). There is also concern that bed bugs may transmit zoonotic viruses from populations of birds and bats to humans (Adelman et al. 2011).

In addition to the danger of pathogen transmission, there is also physical trauma from the bites themselves (Delaunay et al. 2011; Dogget et al. 2012; Williams and Willis 2012; Zorilla-Vaca et al. 2015). Bites most frequently occur on parts of the body uncovered during sleep (Delaunay et al. 2011; Williams and Willis 2012). Wheals from bite lesions are typically 5-20 mm, and itchy, though the severity of the reaction depends on the individual bitten (Dogget et al. 2012; Zorilla-Vaca et al. 2015). Dermatological effects can include dermatitis, ecthyma, folliculitis, impetigo and lymphangitis (Zorilla-Vaca et al. 2015). Some people suffer from allergic reactions due to bed bug saliva, causing red inflammation and triggering anaphylaxis in certain cases (Zorilla-Vaca et al. 2015). If an infestation is heavy, bed bugs can cause anemia and other health problems due to the amount of blood loss they cause the host (Delaunay et al. 2011).

Bed bug infestations and bites can also cause a significant psychological stress (Delaunay et al. 2011; Romero 2011a; Williams and Willis 2012; Zorilla-Vaca et al. 2015). People have been reported to constantly worry and feel insecure, particularly if bitten on the face or neck (Williams and Willis 2012). Formication has also been recorded in people, even after infestations have been exterminated (Williams and Willis 2012). The social stigma attached to bed bug infestations is another dimension of the psychological trauma (Romero 2011a; Williams and Willis 2012). It is erroneously believed that infestations are due to poor hygiene, causing people in afflicted homes to feel shame (Delaunay et al. 2011; Romero 2011a; Williams and Willis 2012; Zorilla-Vaca et al. 2015). The discomfort caused by the bites and infestation can lead to insomnia, further adding to the stress (Romero 2011a; Williams and Willis 2012).

Bed bugs are pests of economic importance (Dogget et al. 2012). They commonly infest poultry farms, causing stress to the chickens and leaving blood and fecal spots on eggs, which reduces their economic value (Szalanski et al. 2008). The hospitality industry is especially affected, with bed bug infestations causing bad publicity and requiring expensive extermination procedures (Delaunay et al. 2011). A very conservative estimate of the economic damage caused by bed bug infestations in the U.S.A. is still well over 3 billion dollars as of the year 2012 (Dogget et al. 2012).

1.3 Global resurgence of *Cimex lectularius*

Before the 1940s, bed bugs were near ubiquitous and generally considered an accepted pest species (Szalanski et al. 2008). It was difficult to treat infestations due to lack of safely applied effective insecticides. In 1930, London had a third of its population affected from bed bug infestations (Usinger 1966; Dogget et al. 2012). Despite a decline in bed bug

numbers in the 1940s and 1950s, they have made a cosmopolitan resurgence in the last twenty years (Reinhardt and Siva-Jothy 2007). Reports of an increase in the number, distribution and intensity of infestations have been reported in Africa, Australia, Canada, Europe, Japan, Malaysia, South Korea and Thailand (Davies et al. 2012).

To date, no formal scientific studies have been conducted on the prevalence of infestations, but many independent reports have been observed that all point to vast increases (Dogget et al. 2012). For example, in Berlin, there were 5 reported cases of infestation in private homes during 1993, while 76 cases were reported in 2004 (Davies et al. 2012). A 2006 survey of 121 pest managers in Australia revealed a 4500% increase of bed bug treatments from 2000 to 2006 (Dogget 2007). The London School of Hygiene and Tropical Medicine surveyed local pest control professionals and discovered the number of bed bug treatments rose an average of 24.7% every year from 2000 to 2006 (Richards et al. 2009). A survey conducted in 2010 by the University of Kentucky and the National Pest Management Association found over 95% of 1000 pest management companies that operate within U.S.A. or internationally reported treating bed bug infestations that year (Davies et al. 2012). In contrast, only 25% of American respondents reported encountering infestations before 2000. Complaints about bed bugs to the New York City council increased from 537 in 2004 to 10,985 in 2009 (Dogget et al. 2012). Notwithstanding the shortcomings of these reports, they likely constitute a somewhat accurate depiction of the rise in bed bug infestations across the world (Dogget et al. 2012).

One of the probable reasons for this resurgence is the increase in human global travel and the use of second hand furniture (Delaunay et al. 2011; Dogget et al. 2012). Additionally, infestation hot spots reported by pest control professionals are often locations

where humans are highly transient (Dogget 2007; Potter et al. 2010). Some of the major locations for infestations include motels, hostels, trains, resorts, hospitals, homeless shelters, public transportation, and movie theatres (Dogget 2007; Potter et al. 2010).

Another proposed reason for the increased prevalence of bed bug infestations is the shift from carbamate and organophosphate insecticides to pyrethroid insecticides, along with the subsequent evolution of pyrethroid resistance in bed bug populations (Romero 2011b; Dogget et al. 2012). Two amino acid substitution mutations that grant pyrethroid resistance due to reduced voltage gated sodium ion channels target-site sensitivity were found in bed bug populations from 17 states across the U.S.A (Zhu et al. 2010). Populations demonstrating insecticide resistance through metabolic detoxification have also been identified (Romero et al. 2011b, Adelman et al. 2011).

A third reason for the global resurgence of bed bug infestations is likely an ignorance of how to detect and exterminate bed bugs (Dogget et al. 2012). During the 50 years following their near extirpation, it was rare for entomologists and pest control professionals in the developed world to encounter a bed bug infestation even once (Davies et al. 2012). Roughly half of the people living in infested homes are unaware they have bed bugs (Wang et al. 2010). Ignorance can also cause unintentional passive dispersal, such as taking used furniture or clothing containing bed bugs into private homes, or lack of inspections of accommodations when travelling and luggage before returning home (Reinhardt and Siva-Jothy 2007; Dogget et al. 2012). Furthermore, if extermination techniques are poorly applied they don't kill the entire population and can cause bed bugs to disperse (Dogget et al. 2012).

1.4 Population dynamics and genetics of *Cimex lectularius*

1.4.1 Population composition

There are limited studies on the infestation dynamics of bed bugs (Wang et al. 2010). In the study of Wang et al. (2010), which was conducted on a high rise apartment complex in Indianapolis, 45% of the units were infested. It was found that 78% of the gathered bugs were nymphs. A different study found that the percentage of nymphs was even higher, accounting for 93% of the captured bugs (Schaafsma et al. 2012). There is evidence of populations having an equal ratio of male to female bugs (Schaafsma et al. 2012). It is estimated that established bed bug populations double in size every 13 days (Polanco et al. 2011).

1.4.2 Dispersal

Bed bugs have been shown to disperse both passively and actively (Pfiester et al. 2009; Wang et al. 2010; Dogget et al. 2012). Passive dispersal is facilitated by humans inadvertently transporting them to new locations on or in their clothing, luggage, or used furniture (Delaunay et al. 2011; Dogget et al. 2012). Evidence for human facilitated passive dispersal is increased interceptions of bed bugs found in luggage by national quarantine inspectors in Australia (Davies et al. 2012). Active dispersal of bed bugs most commonly involves individuals walking from one apartment unit to an adjacent apartment unit, and adult bed bugs have been found in doorways and hallways (Pfiester et al. 2009; Wang et al. 2010). Female adult bed bugs are more frequently found dispersing from aggregations, suggesting they are the most likely dispersal stage (Pfiester et al. 2009; Wang et al. 2010). Of the bed bugs displaying active dispersal, 90% of them were adults (Wang et al. 2010).

1.4.3 Population genetics

Population genetics studies on bed bugs have been completed using nuclear rRNA, mt DNA genes, and microsatellite loci as markers (Szalanski et al. 2008; Vargo et al. 2011; Balvin et al. 2012; Booth et al. 2012; Davies et al. 2012; Booth et al. 2015). The aim of the first of these studies (Szalanski et al. 2008) shed light on the dispersal patterns of bed bugs during their recent global resurgence. Szalanski et al. (2008) focused on genetic variation among various bed bug populations in U.S.A. Canada, and Australia. They examined a partial sequence of the mitochondrial (mt) 16S rRNA gene and nuclear rRNA ITS-1 region of 136 adult bed bugs sampled from 22 populations and found a relatively high genetic diversity in the 16S gene, and low diversity in the ITS-1 gene (Szalanski et al. 2008). Research has also investigated the genetic diversity of microsatellites and the mt COI and 16S genes between dozens of bed bug populations found parasitizing humans versus bed bug populations found parasitizing bats (Balvin et al. 2012; Booth et al. 2015).

The perceived near extirpation of bed bugs from many areas around the world suggests a genetic bottleneck would have occurred, which would be reflected in low genetic diversity across current bed bug populations (Szalanski et al. 2008; Vargo et al. 2011; Balvin et al. 2012; Booth et al. 2012; Davies et al. 2012; Fountain et al. 2014; Booth et al. 2015). However, all of the studies completed thus far have found a relatively high genetic diversity between populations in different locations (Szalanski et al. 2008; Vargo et al. 2011; Balvin et al. 2012; Booth et al. 2012; Davies et al. 2012; Fountain et al. 2014; Booth et al. 2015). Such high diversity across populations is atypical for species that have undergone a recent and single founder event (Szalanski et al. 2008). This evidence suggests that present North American bed bug populations were introduced from areas of the world where there was no

genetic bottleneck, or their diversity was conserved through isolation from human hosts, instead parasitizing birds or bats (Szalanski et al. 2008; Vargo et al. 2011). The latter explanation is bolstered by the common occurrence of bed bugs on chickens in poultry farms and bats (Szalanski et al. 2008; Booth et al. 2015).

Conversely, though there is little evidence for a single founder event across bed bug populations, each population found at a particular locality appears to be the result of a severe founder event (Fountain et al. 2014). All of the studies on bed bugs thus far have found genetic diversity within human associated population to be low, resulting from a great deal of inbreeding (Szalanski et al. 2008; Vargo et al. 2011; Balvin et al. 2012; Booth et al. 2012; Davies et al. 2012; Fountain et al. 2014; Booth et al. 2015). There is a consensus among research that a small propagule (a single gravid female) is sufficient to infest an entire apartment, and most subpopulations found within a locality are from the same lineage (Szalanski et al. 2008; Vargo et al. 2011; Balvin et al. 2012; Booth et al. 2012; Davies et al. 2012; Fountain et al. 2014; Booth et al. 2015). It is thought that repeated extermination events leads to frequent local extirpation and colonization. These frequent founder events are hypothesized to select for a resistance to the deleterious effects of inbreeding (Fountain et al. 2014). The discrete and repeated founder effects, coupled with limited migration, contribute to strong genetic differentiation between populations (Fountain et al. 2014).

Genetic diversity was found to be even greater among populations associated with bat hosts (Balvin et al. 2012; Booth et al. 2015). It was also found that the levels of relatedness within populations located in bat roosts were much lower than human-associated populations (Balvin et al. 2012; Booth et al. 2015). These same studies provide evidence to suggest host

driven lineage divergence in *C. lectularius*, leading to host races (Balvin et al. 2012, Booth et al. 2015).

1.5 *Cimex adjunctus* and *Oeciacus vicarius*

The eastern bat bug, *C. adjunctus*, commonly feeds on bats, but is capable of feeding on humans when its preferred host is unavailable (Usinger 1966). *Cimex adjunctus* is not known to directly vector any disease, but a species of *Bartonella* was detected in a *C. adjunctus* specimen in Florida using PCR (Reeves et al. 2005). Like the common bed bug, the eastern bat bug falls within the subfamily Cimicinae (Usinger 1966).

The swallow bug, *Oeciacus vicarius*, commonly feeds on nesting cliff swallows and house sparrows, but can also feed on other birds or even mammals (Brown and Brown 2005). *Oeciacus vicarius* is not thought to vector any pathogenic agents to humans, though it is the main arthropod vector of a recombinant alphavirus known as Buggy Creek virus, which affects birds (Brown and Brown 2005). Like *C. lectularius*, *O. vicarius* falls within family Cimicidae and subfamily Cimicinae (Usinger 1966).

1.6 Systematics of the family Cimicidae

There is a paucity of data and analysis regarding phylogenetic relationships below infraorder Cimicomorpha (Schuh et al. 2009; Kocher et al. 2015). The proposed evolutionary relationships of the taxa within superfamily Cimicoidea and the family Cimicidae are based on morphological characters, chromosome numbers, and host associations (Usinger 1966, Balvin et al. 2015). Previous studies have associated the Cimicidae with other families within the superfamily Cimicoidea using both morphological and molecular characters (Schuh et al. 2009; Cui et al. 2013). A total evidence analysis using 16S, 18S, 28S and COI DNA sequence data and 73 morphological characters (Schuh et al. 2009) has determined

infraorder Cimicomorpha and superfamily Cimicoidea are both monophyletic. This same study reported that the families Cimicidae, Polyctenidae and Curaliidae form a monophyletic clade (Schuh et al. 2009). It has been postulated, based on comparisons of the DNA sequences of mitochondrial genomes, that the divergence of superfamily Cimicoidea from Naboidea occurred 161 mya (Jung and Lee 2012). However, no DNA sequences of the mitogenome of any species within species within Cimicidae were included in this study (Jung and Lee 2012).

Similarly, the proposed relationships of the species within the family Cimicidae, which are subdivided into 6 subfamilies and 24 genera (Henry 2009), are inferred from morphological characters, chromosome numbers, and host associations (Balvin et al. 2015). However, there is little evidence that these characters provide suitable phylogenetic resolution (Balvin et al. 2015). Comparisons of the mitogenomes of *C. lectularius, C. adjunctus* and *O. vicarius* may provide the characters required to resolve phylogenetic relationships within family Cimicidae (Schuh et al. 2009; Balvin et al. 2015).

1.7. Mitochondrial DNA and mitogenomics

The mitochondrial (mt) genomes of insects are usually a single circular, double stranded molecule with a length that ranges from 14-20 kb (Li et al. 2011; Kocher et al. 2015). They typically contain 13 protein-coding genes (PCGs), two ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes and a non-coding AT-rich region known as the control region (CR), which plays a part in initiation of replication and transcription (Li et al. 2011; Li et al. 2012a; Li et al. 2012b; Kocher et al. 2015). The CRs of different insect species have been shown to differ in their sequence, organization, and location in relation to the genes, suggesting it is possible they may not be homologous between different taxa (Li et al. 2011;

Li et al. 2012a; Li et al. 2012b). Additionally, the length of CR's are highly variable due to their high rate of substitutions and variable number of tandem repeats (Li et al. 2011; Li et al. 2012a; Li et al. 2012b). Most insect mt genomes have been compared to the mitogenome of *Drosophilia yakuba* (Li et al. 2011; Li et al. 2012a; Li et al. 2012b). This fly was the first insect to have its complete mt genome sequenced (Clary and Wolstenholme 1985). The order of the genes in the mt genome of *D. yakuba* is shared by most insect species, and is considered to represent the "ancestral" gene arrangement for the class Insecta (Li et al. 2011; Li et al. 2012a; Li et al. 2012b).

Studies of insect mitochondrial genomes outnumber nuclear genomes ten to one and whole mitochondrial genomes have been sequenced for species in every insect order (Cameron 2014a; Kocher et al. 2015). The mitochondrial DNA is used because it is plentiful in animal tissues, the genome is small compared to the nuclear genome, it has a fast rate of evolution, and it is inherited maternally, with a low likelihood of recombination (Li et al. 2011; Li et al. 2012a; Li et al. 2012b; Cameron 2014a; Kocher et al. 2015).

Insect mitogenomes are most commonly used in phylogenetic analyses, with well over 100 journal articles using this data to build phylogenies (Cameron 2014a). The taxonomic levels these analyses cover ranges from individual species phylogeographic studies to interclass studies, and every taxonomic rank in between (Cameron 2014a). While some phylogenetic studies only focus on particular genes, it has been argued that as many genes as possible should be included to eliminate potential bias (Cameron 2014a).

1.8 Objectives of this thesis

Over 50 mitogenomes have been completely sequenced in suborder Heteroptera, covering 35 different families (Kocher et al. 2015). Despite this, there are at this stage, no published

details on the complete mitogenomes of any species within the family Cimicidae. The aim of my MSc research is to determine the DNA sequences of the mitochondrial genomes of *C. lectularius, C. adjunctus, and O. vicarius*. Sequencing the complete mt genomes of species within the Cimicidae has implications for primer design, population genetics and gene flow, as well as testing hypotheses relating to evolutionary history and phylogeny (Li et al. 2011; Li et al. 2012a; Li et al. 2012b; Cameron 2014; Kocher et al. 2015).

Molecular markers useful for understanding population genetics and dispersal of cimicids can be found in mt DNA (Li et al. 2011; Li et al. 2012a; Li et al. 2012b; Yang et al. 2013; Booth et al. 2015). Knowledge of the genetic variation within populations of medically-important insect species provides a basis for a better understanding of insecticide resistance, infestation dynamics and dispersal patterns (Szalanski et al. 2008; Vargo et al. 2011; Saenz et al. 2012; Yang et al. 2013; Cameron 2014a; Kocher et al. 2015). Sequence data from the 16S and COI mt genes have been used for such studies in *C. lectularius*, and finding other gene regions that have intraspecific variation between populations will add to the resolution of these studies (Li et al. 2011; Li et al. 2012a; Li et al. 2012b; Yang et al. 2013; Booth et al. 2015). The mt DNA sequence of the three will be examined for regions that display intraspecific variation between different populations.

Comparisons of the DNA sequences of the complete mitogenomes of organisms can also be useful to inform evolutionary relationships (Li et al. 2011; Li et al. 2012a; Li et al. 2012b; Cameron 2014a; Kocher et al. 2015). Recent studies attempted to discern the infraordinal relationships of Heteroptera, yet none have included a member of Cimicidae (Li et al. 2011; Li et al. 2012a; Li et al. 2012b). Furthermore, phylogenetic relationships within Cimicidae are currently based on the paragenital system, host relations, bristle shape, and

chromosome pattern (Reinhardt and Siva-Jothy 2007). Determination of the DNA sequences of *C. lectularius*, *C. adjunctus* and *O. vicarius* will provide the opportunity to test hypotheses regarding the phylogenetic relationships of cimicids to other heteropterans (Li et al. 2011; Li et al. 2012a; Li et al. 2012b; Cameron 2014a; Kocher et al. 2015).

2. MATERIALS AND METHODS

2.1 *Cimex lectularius*

2.1.1 Collection and identification of specimens

A total of 36 living adult specimens (Cim1 to Cim30 and Cl1-Cl6) of the common bed bug were collected from four homes in Saskatoon (Table 2.1). Individual samples were gathered using forceps, placed in 1.5mm microcentrifuge tubes, and fixed in a 70% ethanol. Each bed bug was identified using morphological features and the taxonomic key of Usinger (1966). Diagnostically useful features included rostrum length, mesothoracic coxae separation, antennal segment length, as well as pronotal anterior margin and hairs (Figures 2.1 and 2.2). The species identity of all 36 bed bugs was confirmed by performing PCR of the COI and 16S gene regions [see section 2.1.4]. Amplicons were sequenced and compared with known sequences of *C. lectularius* on NCBI's GenBank using BLAST.

TABLE 2.1
Collection locations and dates of *Cimex lectularius* in Saskatoon (2012)

Sample Code for Specimens	Date Collected	Location
Cim1 to Cim10	May 14th	Priel Cr., Saskatoon, SK
Cim11 to Cim20	May 23rd	Ave. P North, Saskatoon, SK
Cim21 to Cim30	June 7th	Pendygrasse Road, Saskatoon, SK
Cl1 to Cl6	Nov. 2nd	Preston Ave., Saskatoon, SK

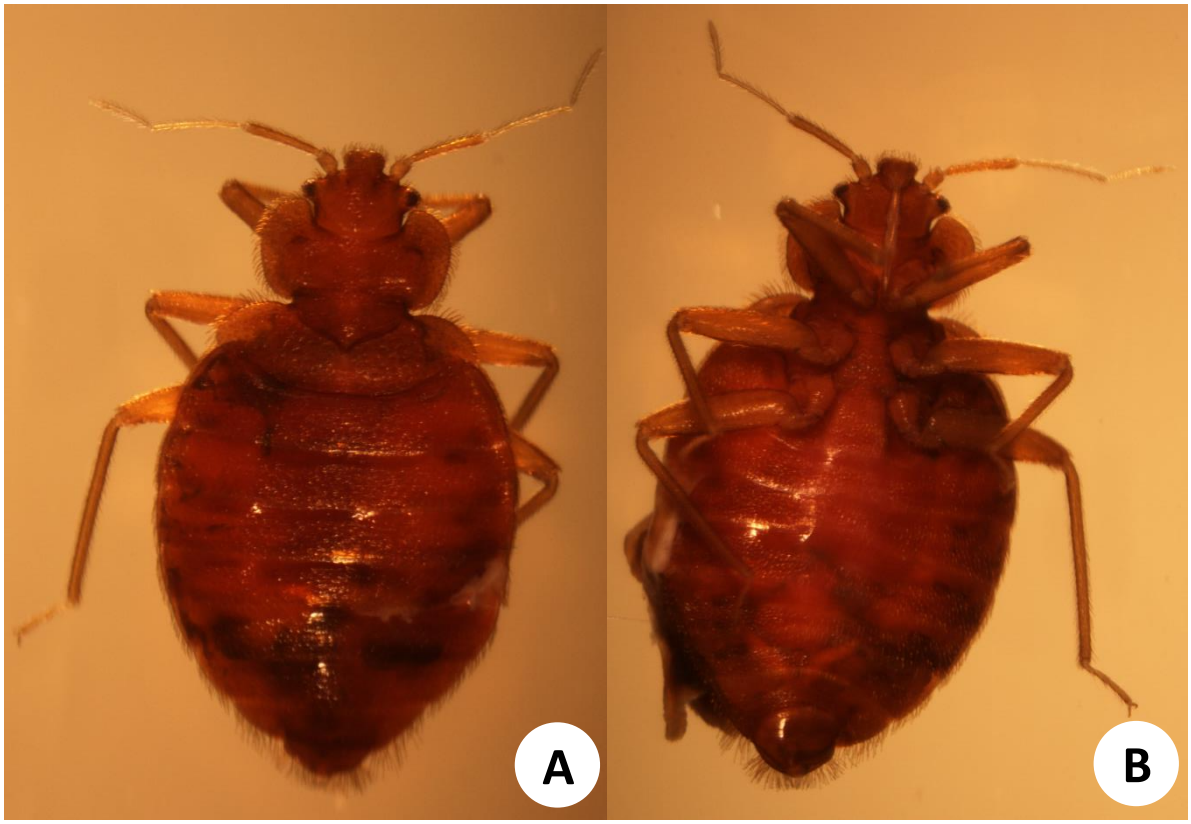


Figure 2.1. Dorsal (A) and ventral (B) view of a male *Cimex lectularius* specimen.

Figure 2.2. The anterior end of a male *C. lectularius* specimen showing the pronotal hairs (A) and the relative lengths of the third and fourth antennomeres (B) characters, which were morphological traits used to differentiate among *C. lectularius, C. adjunctus,* and *O. vicarius.*

2.1.2 DNA extraction

Each bed bug was transferred to an individual 1.5 ml micropestle tube to which 180 µl of ATL buffer from a DNeasy® Blood and Tissue Kit (QIAGEN® Sciences, Maryland, USA) was added. Bed bug bodies were ground up using disposable micropestles and a cordless drill. A total of 20µl of proteinase K (QIAGEN®) was added to each tube after the tissues and exoskeleton were successfully fragmented. Samples were then mixed using a vortex and incubated at 55°C for 24 hours. The samples were thoroughly mixed using a vortex again and a 200 µl volume of AL buffer (QIAGEN®) was added to each tube prior to mixing their contents with a vortex and incubating them at 70°C for 10 minutes. A total of 200 µl of 100% ethanol was added to each sample which was then mixed using a vortex. Each mixture was pipetted into separate DNeasy Mini spin columns that were placed in 2ml collection tubes. Samples were centrifuged for 1 minute at 6,000 RCF. The flow-through and collection tubes were discarded, while the spin columns were placed in new 2ml collection tubes. A total of 500 µl of AW1 buffer (QIAGEN®) was added to the each spin column prior to centrifugaton at 6,000 RCF for 1 minute. The second flow-through and collection tubes were also discarded and replaced by new collection tubes. A total of 500 µl of AW2 buffer (QIAGEN®) was then added to each spin column before being centrifuged at 18,000 RCF for 3 minutes to dry the DNeasy membranes. The flow-through and collection tubes were discarded and the DNeasy Mini spin columns were placed in clean 1.5 ml microcentrifuge tubes. Fifty µl of AE buffer (QIAGEN®) was placed directly on each DNeasy membrane and incubated at room temperature for 5 minutes. The spin columns were then centrifuged at 6,000 RCF for 1 minute to elute the DNA. This elution process was then repeated using another 50 µl of AE buffer (QIAGEN®) resulting in 100 µl of pooled eluates. The concentration of genomic

DNA solutions were determined by spectrophotometric analysis using a NanoDrop® system (Table 2.2). Genomic DNA was stored at -20°C in 1.5 ml microcentrifuge tubes.

TABLE 2.2
Spectrophotometric analysis of *Cimex lectularius* DNA templates

Sample Code	260/280 Ratio[a]	Concentration (ng/µl)[b]
Cim6	2.56	20.2
Cim15	2.22	10.0
Cim17	2.06	17.0
Cim26	1.90	11.9

a: 260/280 is the ratio of absorbance at 260 and 280 nm, and is indicative of the purity of nucleic acids. For DNA, values falling below 1.8 can indicate contamination (NanoDrop®).
b: The concentration is the nanograms of DNA template per µl of AE buffer

2.1.3 Molecular identification of bed bugs

The species identity of individual bed bugs was determined by amplification of two mitochondrial genes, COI and the 16S rRNA gene. This also permitted an initial assessment of the approximate quality of the genomic DNA extracted from individual bedbugs. Primers LR-J-13007 and LR-N-13398 (Tables 2.3 and 2.4) were used to amplify 428 bp of the 16S rRNA gene in *C. lectularius* using the amplification conditions described by Szalanski et al. (2008). In addition, primers LEP-F and LEP-R (Tables 2.3 and 2.4) were used to amplify 658 bp of the COI gene in *C. lectularius* using the amplification conditions described by Balvin et al. (2012). Amplicons were subjected to agarose gel electrophoresis. Agarose (1.5%) gels were made using 0.75g of agarose powder (Invitrogen), 50ml of 1X TBE buffer (OmniPur®), and 2.5µl SYBR Safe dye. Agarose gels were run at 120 volts for 40 minutes. Amplicons were visualized using UV light transillumination. The size of amplicons was determined using the Thermo Scientific GeneRuler 100 bp Plus DNA Ladder (1 µg/µl). Samples were purified using the ExoSAP-IT® approach, an enzymatic elimination of unincorporated primers and dNTPs. They were then subjected to automated DNA sequencing using both forward and reverse primers in separate reactions. DNA sequences were compared with those of *C. lectularius* on GenBank using the program BLAST.

2.1.4 Amplification and characterization of the Mt genome

Attempts were made to amplify the entire mitochondrial genome sequence of *C. lectularius* in two overlapping fragments using a long PCR technique (Cheng et al. 1994; Hu et al. 2002) (Figure 2.3). The initial long range PCRs performed on the bed bugs were conducted using the Roche Expand Long Template PCR System and their suggested protocol (Roche, REF 11681842001). Long PCRs were attempted using the primer pairs LR-J-13007 and LEP-R,

LEP-F and LR-N-13398, mtBUG-16S-F and mtBUG-COX-R, and mtBUG-COXF and mtBUG-16S-R (Tables 2.3 and 2.4). Despite troubleshooting each of these reactions, no DNA fragment was successfully amplified using this kit. Subsequent attempts at long PCRs were made with the Clontech Advantage® 2 PCR Enzyme System, using the protocol and thermal cycling conditions recommended in the user manual. Eight primers with high GC content (i.e., mtBUG-COXF, mtBUG-COX-R, LONG F1, LONG F2, LONG R1, LONG R2, Cimex-LMF1, and Cimex-LMR1; Tables 2.3 and 2.4) were designed to work with this protocol. Amplification reactions were completed using the primer pairs LONG F1 and Cimex-LMR1, and Cimex-LMF1 and LONG R2, though they each yield shorter fragments than expected (Table 2.3, Table 2.4, Figure 2.4).

Additional primers were designed by obtaining the complete mtDNA sequence of species in the superfamily Cimicoidea hypothesized to be closely related to *C. lectularius* (Weirauch and Schuh 2011; Li et al. 2013) (Table 2.5). These sequences were compared with each other, looking for regions that were identical or similar across many species. Further primers were designed by comparing expressed sequence tag (EST) data of *C. lectularius* to the complete mtDNA sequence of *Orius niger* (hereafter referred to as *Or.niger*), looking for identical regions (Hua et al. 2008; Francischetti et al. 2010; Moriyama et al. 2012) (Figure 2.4). *Orius niger* was chosen as it is hypothesized to be the closest relative to *C. lectularius* to have its entire mitochondrial DNA sequence available on Genbank (Hua et al. 2008; Weirauch and Schuh 2011; Li et al. 2013). All of the primers that were successfully used to amplify mtDNA from *C. lectularius* are shown in Tables 2.3 and Table 2.4, and in Figure 2.4.

The short range PCRs performed on the bed bugs were initially conducted using either Taq DNA polymerase and reagents (Fermentas, Massachusetts, USA), or iTaq™ DNA

polymerase and reagents (Bio-Rad Laboritories, California, USA.). Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific Inc. 2011) was used for difficult to amplify regions. Initial concentrations of the PCR reagents for each reaction were 2 mM Mg++, 1 µM for each primer, and 200 µM for each dNTP. Initial thermal cycling conditions for each PCR were 30 cycles of 94°C for 45 seconds of denaturation, 48-58°C (depending on the T_m of the primers) for 45 seconds of primer annealing, and 72°C for 60 seconds of extension. These 30 cycles were preceded by an initial denaturation of 5 minutes at 94°C and followed by a final extension of 5 minutes at 72°C.

If a PCR reaction did not produce a product, or produced multiple bands when subjected to gel electrophoresis, troubleshooting the PCR involved several steps. First, the PCR was rerun using a temperature gradient during the annealing step, ranging from 46-62°C. If this did not produce a desired product, Mg^{++} concentration varied from 1.4 to 2.6 mM in increments of 0.3 mM, each concentration being tested using a temperature gradient. Lastly, primer concentrations were changed to 0.6 and 0.2 µM and tested using a temperature gradient.

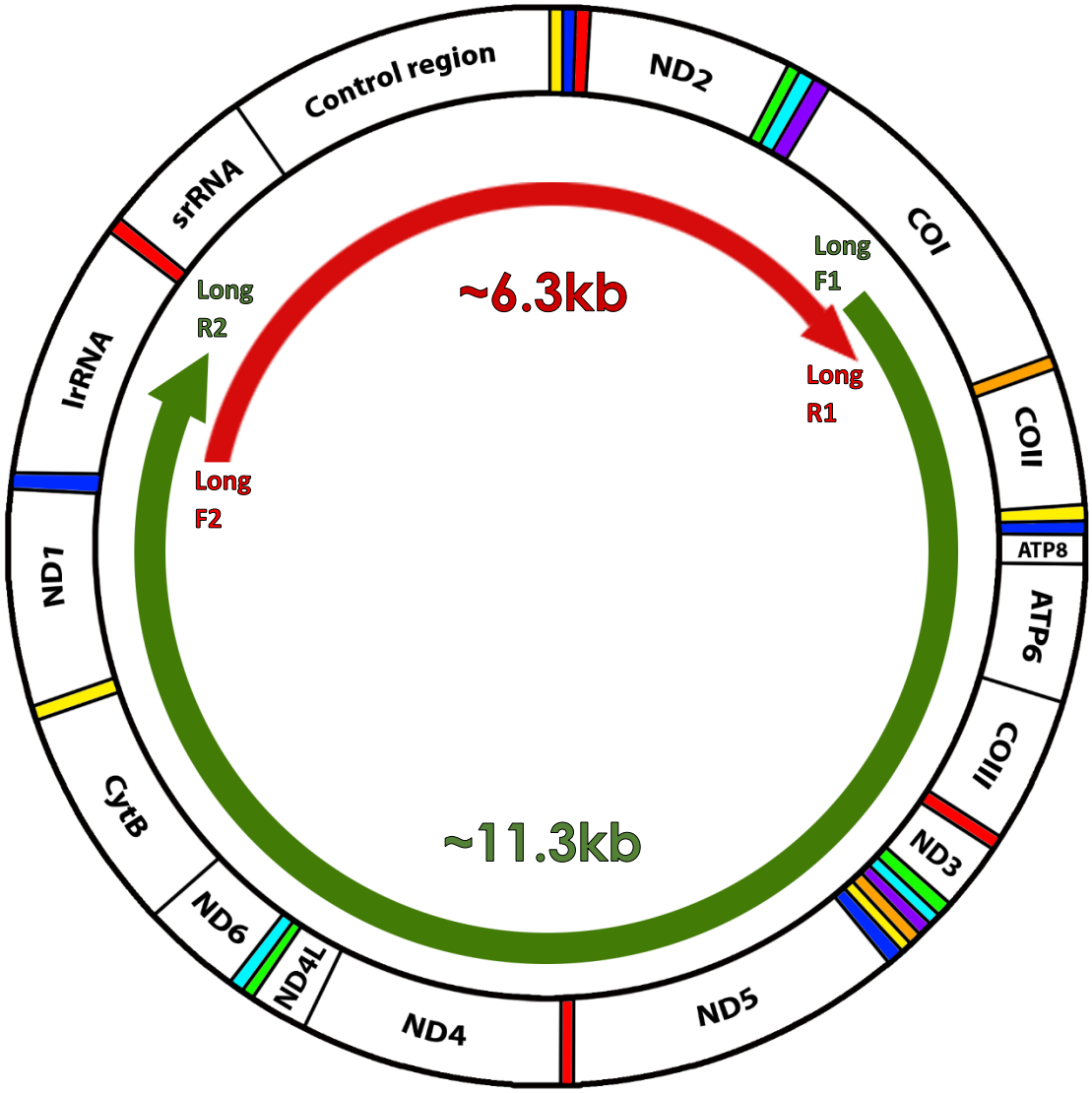


Figure 2.3. Schematic representation of long range PCR. Amplicons of *C. lectularius* mtDNA from the Long F1 and Long R2 primer set would be roughly 11.3 kb, and the Long F2 and Long R1 primer amplicons would be roughly 6.3 kb, encompassing the entire mitogenome between the two overlapping fragments.

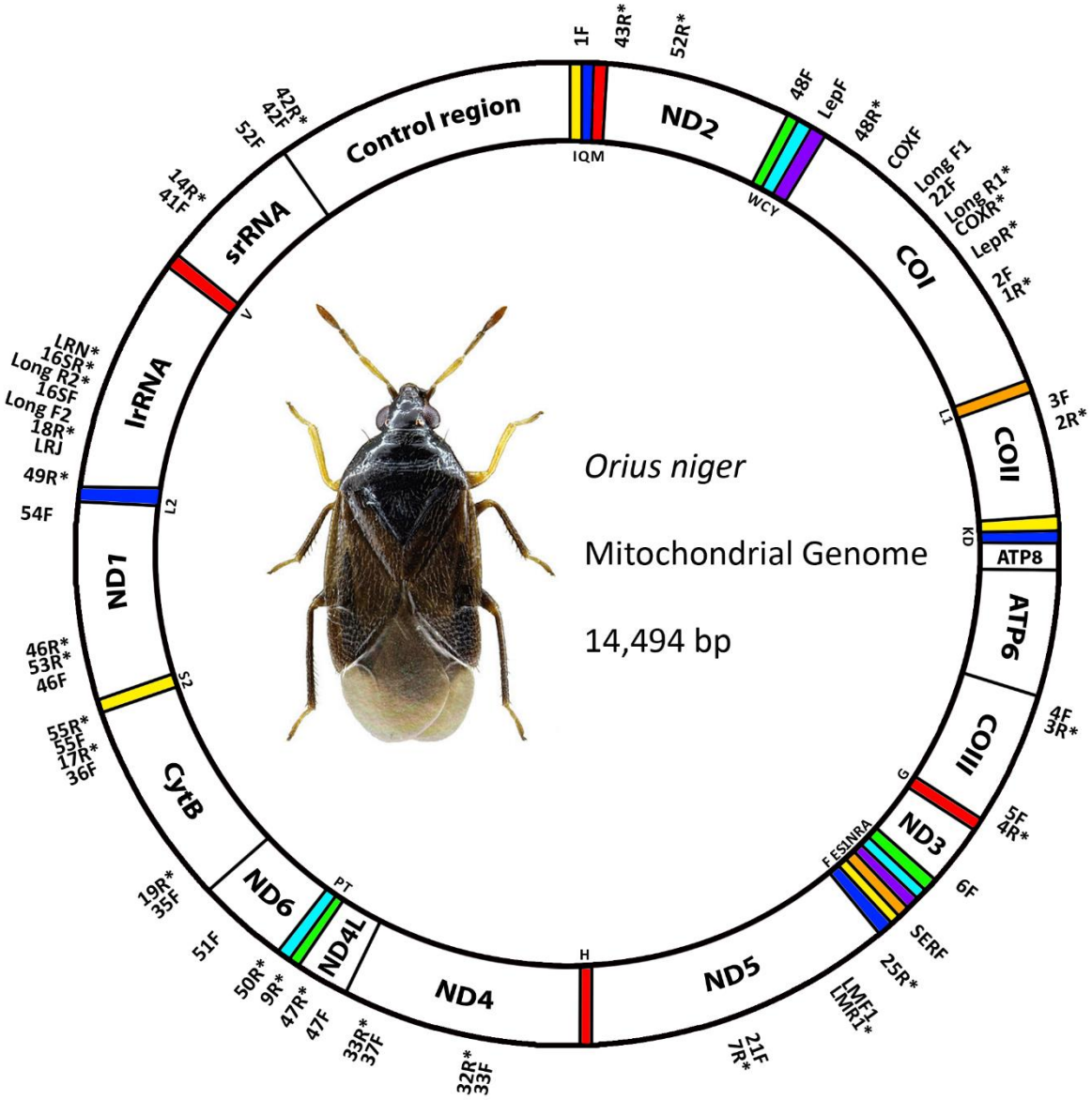


Figure 2.4. Location of PCR primers used to amplify the mtDNA of *Cimex lectularius* relative to the mitogenome of *Orius niger*. Reverse primers are signified with an asterisk, tRNAs are labelled according to IUPAC-IUB single-letter amino acid codes. Image of *Or. niger* from Raupach et al. (2014).

TABLE 2.3
Forward PCR primers used to amplify the mtDNA of *C. lectularius*

Primer	Sequence	Position*
Bbug-Mt-1F	AGCTATTAGGTTCATACCCTA	149
Cimex-NC-48F	AACTATTAACCTTCAAAGTTA	1,185
LEP-F	ATTCAACCAATCATAAAGATATNGG	1,316
mtBUG-COXF	ATTGGGGGATTTGGGAATTGACTAGTGCC	1,570
Long F1	TGATAATTGGCGCACCTGATATAGCATTC	1,601
Bbug-Mt-22F	ATAAGATTTTGACTTTTACCTCC	1,645
Cimex-NC-2F	ATTTACCATTGGCGGTATAACAGG	2,388
Cimex-NC-3F	ACGCCAGCAGTAACATTATTATTC	3,151
Cimex-NC-4F	ACCACCCATATCATATAGTAGATTA	4,610
Cimex-NC-5F	AGCCTCATGATACTGACACTTC	5,301
Cimex-NC-6F	ACTGCTGGGATTATATTACGAAT	5,740
SER-F	GCTAACTATCTTTTAAAGCGG	6,021
Cimex-LMF1	ACTTTGACAGCAATATAAACTCTCTCTGCC	6,649
Bbug-Mt-21F	CAGCAGTTACTAAAGTAGATGAG	7,220
Cimex-NC-33F	ACCATAGCTAAATTTCTTCAACT	8,144
Cimex-NC-37F	AACAGGAGCCTCAACATGAGC	8,648
Cimex-NC-47F	ATAACCAGATCCTAACCCTG	9,394
Cimex-NC-51F	ATCAAATGTGGCCTCTAATG	10,070
Cimex-NC-35F	ATCCTTACTAAGAATATGTTTGT	10,276
Cimex-NC-36F	AACCTCAGCTTCTAGGAGATCC	10,912
Cimex-NC-55F	ATGAATTGGCGCTAAACCAGCC	11,209
Cimex-NC-46F	AGTTTATCATACCGATACCGG	11,500
Cimex-NC-54F	ACAGGGATATCTTTGTTCCTTA	12,142
LR-J-13007	TTACGCTGTTATCCCTAA	12,612
LongF2	AATAAATTTACCTGTCGCCCCAACAAAACT	12,703
mtBUG-16S-F	GTCAATCCCTCGTATGACCATTCATACTAGTTTCC	12,872
Cimex-NC-41F	ACATTCAATAACAGAGGTATAC	14,029
Cimex-NC-52F	GAGCATTATTAGTCCCCTCTCG	14,500
Cimex-NC-42F	AACTATATCCTTATTCATTAGTT	14,661

***** The nucleotide position listed is the approximate position in relation to the mitochondrial DNA sequence of *Orius niger*.

TABLE 2.4
Reverse PCR Primers used to amplify the mtDNA of *C. lectularius*

Primer	Sequence	Position*
Cimex-NC-43R	ATCTTAATACTAGTAAAGAACC	247
Cimex-NC-52R	CATCCTTCTGAGAGTGATCTG	350
Cimex-NC-48R	AGTATACCTGCTCATATCCC	1,414
Long R1	TAAGGTTATCCCCGCAGGACGTATATTTAA	1,851
mtBUG-COX-R	ATAAGGGAGTTCGTTCTAAGGTTATCCCCGCAGG	1,856
LEP-R	TAAACTTCTGGATGTCCAAAAAATCA	2,053
Cimex-NC-1R	ACGTAATAAGTATCATGTAATGC	2,443
Cimex-NC-2R	AATCAGAATATTCGTAGCTTC	3,265
Cimex-NC-3R	GTTAATGTTAATGCTCCAATTG	4,658
Cimex-NC-4R	ATCTTCCTCATCAATAAATGATA	5,352
Bbug-Mt-25R	AGGGTTTCAGTTAATGTTTCAGG	6,259
Cimex-LMR1	TGGCAGAGAGAGTTTATATTGCTGTCAAAG	6,649
Bbug-Mt-7R	GCTCAAATTCCTTTTTCTTC	7,312
Cimex-NC-32R	TGTTAGGTGAAGTATTACTGTTG	8,179
Cimex-NC-33R	GCTTCTTTACCCCTTTTATTAGG	8,811
Cimex-NC-47R	GGTCTTTTTGCTTATTGTACAGC	9,494
Bbug-Mt-9R	TCTATGATTTACAAGATCAT	9,578
Cimex-NC-50R	ACTAGCTGCCAACAAGTAGT	9,675
Cimex-NC-m19R	GCTAATTCAATGTTAGCTGTAT	10,339
Cimex-NC-m17R	GCAAATAGAAAATATCATTCTGG	10,983
Cimex-NC-55R	ACATAGTTGATAGGGCTTGGCC	11,252
Cimex-NC-53R	ACTTGTATTTTGTTTTGAGGGGC	11,600
Cimex-NC-46R	ATTCTTTTTATAAGAACTTTTAC	11,620
Cimex-NC-49R	GCGTAAGCCAGGTTGGTTTCTA	12,473
Bbug-Mt-18R	AGTTTTGTTGGGGCGACAGGTA	12,694
LONG R2	CGCAGTATTTTAACTGTGCAAAGGTAGCAT	12,926
mtBUG-16S-R	ATGGCCGCAGTACTTTAACTGTGCAAAGG	12,932
LRN	CGCCTGTTTATCAAAAACAT	13,009
Bbug-Mt-14R	TAAAAGAATTTGGCGGTAATT	14,057

***** The nucleotide position listed is the approximate position in relation to the mitochondrial DNA sequence of *Orius niger*.

2.1.5 DNA sequence analyses

Amplicons that were purified as described in section 2.1.2 were submitted to the National Research Council Canada, Plant Biotechnology Institute for automated DNA sequencing using both forward and reverse primers in separate reactions. This sequencing approach produced an approximately 1000 bp sequence from each end of each PCR product submission (Kim and Yeung 1997). Overlapping sequences were assembled with one another to construct the entire mitochondrial genome.

The location and amino acid composition of the protein-coding genes in *C. lectularius* were discerned by inputting regions of the assembled nucleotide sequence into the translate tool offered at ExPASy: Bioinformatics Resource Portal, using the "invertebrate mitochondrial" genetic code setting (http://web.expasy.org/translate/). The amino acid sequences were then entered into NCBI's Protein BLAST tool to confirm which protein-coding gene they represent (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Regions of the assembled nucleotide sequence were then entered into ARWEN, a program that detects tRNA genes in metazoan mtDNA sequence (Laslett and Canbäck 2008). The identity, direction, location, length and secondary structure of the tRNA genes of *C. lectularius* were determined using the output of this program (Laslett and Canbäck 2008). The locations, directions, and lengths of the small ribosomal RNA (srRNA) and large ribosomal RNA (lrRNA) genes were determined by comparing regions of the assembled *C. lectularius* sequence with the known sequence of the srRNA and lrRNA genes of *Or. niger* and other close relatives in Genbank. Their locations were confirmed by their known bordering tRNA genes, which were detected as described above (Laslett and Canbäck 2008).

Phylogenetic analyses were conducted by inputting the amino acid data from *C. lectularius* and twenty-two other heteropterans (Table 2.5) into PAUP* version 4.0b10, a suite of tools for inferring and interpreting phylogenetic trees (Swofford 2002). Both neighbor joining and maximum parsimony trees were deduced using PAUP*. For maximum parsimony analysis, all characters were given equal weight and gaps were treated as amino acid characters. The neighbor joining method used 1000 replicates while the maximum parsimony bootstrap method used 100 replicates.

TABLE 2.5
Sources of mitochondrial amino acid sequences of heteropterans used in phylogenetic analysis

Infraorder	Superfamily	Family	Species	GenBank Accession #
Cimicomorpha	Miroidea	Tingidae	*Corythucha ciliata*	NC_022922
			Pseudocysta persea	KM278221
		Miridae	*Apolygus lucorum*	NC_023083
			Lygus lineolatus	NC_021975
			Adelphocoris fasciaticollis	NC_023796
			Nesidiocorus tenius	NC_022677
	Naboidea	Nabidae	*Himacerus apterus*	JF927831
			Nabis apicalis	JF907590
			Gorpis annulatus	JF907591
			Alleorhynchus bakeri	HM235722
	Reduvoidea	Reduviidae	*Triatoma dimidiata*	NC_002609
			Oncocephalus breviscutum	KC887527
			Valentia hoffmanni	NC_012823
			Pierates arcuatus	KF752445
			Sirthenea flavipes	HQ645959
			Agriosphodrus dohrni	NC_015842
	Cimicoidea	Anthocoridae	*Orius niger*	EU427341
			Orius sauteri	KJ671626
Nepomorpha	Corixoidea	Corixidae	*Sigara septemlineata*	FJ456946
Leptopodomorpha	-	Saldidae	*Sandula arsenjevi*	EU427345
Gerrimorpha	Gerroidea	Gerridae	*Aquarius paludum*	FJ456944

* Classification of taxa following that of Kocher et al. (2010).

2.2 *Cimex adjunctus*

2.2.1 Collection and identification of specimens

Three specimens (Bat1, Bat2 and Ca1) of *Cimex adjunctus* were collected from two different homes (Table 2.6). Individual samples were identified morphologically as described in section 2.1.1. Diagnostically useful features included the rostrum length, mesothoracic coxae separation, antennal segment length, as well as pronotal anterior margin and hairs (Figure 2.5 and 2.2) (Usinger 1966). The DNA of each bat bug was extracted and analyzed as described in section 2.1.2 (Table 2.7). The species identity of each specimen was confirmed using molecular methods as described in section 2.1.4.

TABLE 2.6
Collection locations and dates of *Cimex adjunctus* in 2012

Sample Code for Specimens	Date Collected	Location
Bat1, Bat2	May	Saskatoon, SK
Ca1	July	Saskatoon, SK

TABLE 2.7
Spectrophotometric analysis of *Cimex adjunctus* DNA Templates

Sample Code	260/280 Ratio[a]	Concentration (ng/µl)[b]
Bat1	2.29	15.4
Bat2	2.12	49.7
Ca1	2.15	16.6

a: 260/280 is the ratio of absorbance at 260 and 280 nm, and is indicative of the purity of nucleic acids. For DNA, values falling below 1.8 can indicate contamination (NanoDrop®).
b: The concentration is the nanograms of DNA template per µl of AE buffer

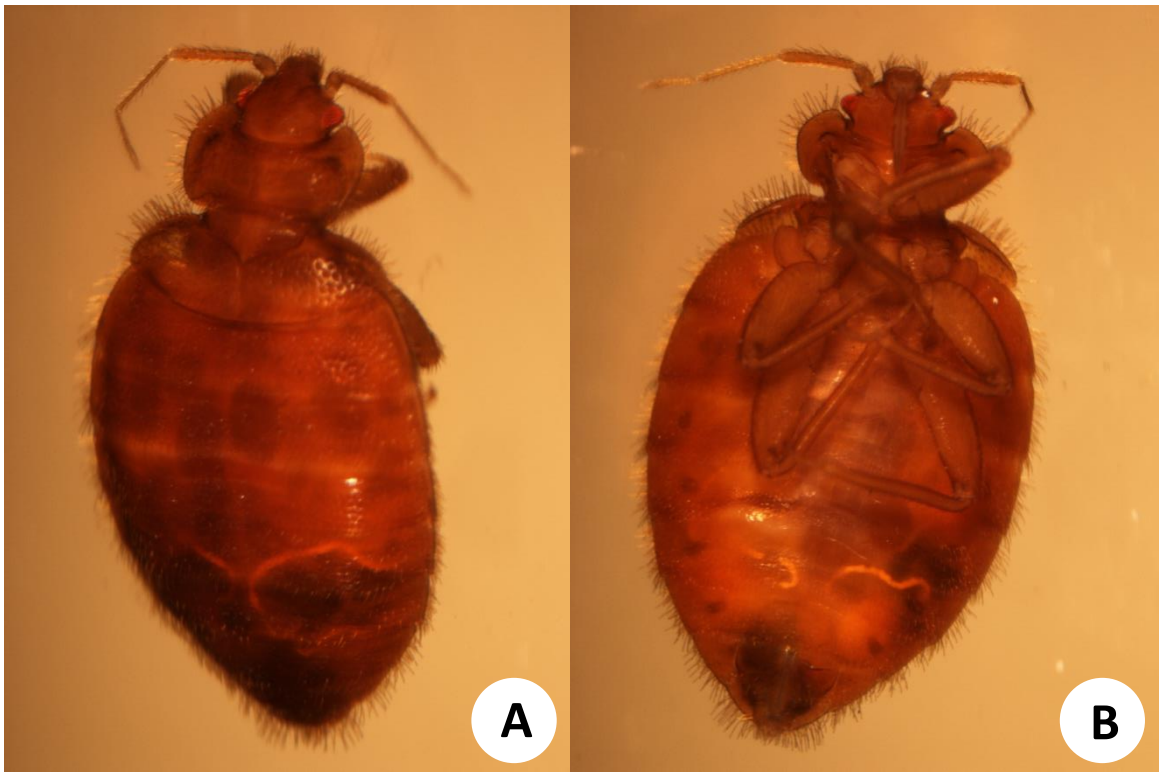


Figure 2.5. Photographs of a male *Cimex adjunctus* specimen showing the dorsal view (A) and ventral view (B).

2.2.2 Amplification and characterization of the mt genome

Attempts were made to amplify the entire mitochondrial genome sequence of *C. adjunctus* in two overlapping fragments as described in section 2.1.5. Despite troubleshooting each of these reactions, no DNA fragment was successfully amplified using either the Roche Expand Long Template PCR System or the Clontech Advantage® 2 PCR Enzyme System. PCR primers were designed to amplify regions of the mt genome of *C. adjunctus* as described in section 2.1.5. All of the primers that were successfully used to amplify mtDNA from *C. adjunctus* are shown on Tables 2.8 and 2.9. Methods and troubleshooting for short range PCRs performed on the bat bugs were conducted as described in section 2.1.5. DNA analyses performed on the sequence data of *C. adjunctus* were performed as described in section 2.1.6.

TABLE 2.8
Forward PCR primers used to amplify the mtDNA of *Cimex adjunctus*

Primer	Sequence	Position*
Bbug-Mt-1F	AGCTATTAGGTTCATACCCTA	149
LEP-F	ATTCAACCAATCATAAAGATATNGG	1,316
Long F1	TGATAATTGGCGCACCTGATATAGCATTC	1,601
BBLong F1	ATGATTGGAGCTCCCGATATAGCATTCCC	1,603
OV-NC-24F	ACTATAATTATCGCAGTCCC	2,270
Cimex-NC-4F	ACCACCCATATCATATAGTAGATTA	4,610
Cimex-NC-33F	ACCATAGCTAAATTTCTTCAACT	8,144
LRJ	TTACGCTGTTATCCCTAA	12,612

***** The nucleotide position listed is the approximate position in relation to the mitochondrial DNA sequence of *Orius niger*.

TABLE 2.9
Reverse PCR primers used to amplify the mtDNA of *Cimex adjunctus*

Primer	Sequence	Position*
BB-NC-48R	AACATGCCTGCTCATATTCCG	1,414
BB-NC-22R	ATGAGCGATATTCCCTGATAGAG	1,737
LEP-R	TAAACTTCTGGATGTCCAAAAAATCA	2,053
Bbug-Mt-24R	CTTGTTAATCCTCCAATTGTAATT	2,407
BB-NC-3R	ATCATACAGTTGTTCCAAAAG	4,672
Cimex-NC-4R	ATCTTCCTCATCAATAAATGATA	5,352
Cimex-NC-33R	GCTTCTTTACCCCTTTTATTAGG	8,811
LRN	CGCCTGTTTATCAAAAACAT	13,009

***** The nucleotide position listed is the approximate position in relation to the mitochondrial DNA sequence of *Orius niger*.

2.3 *Oeciacus vicarius*

2.3.1 Collection and identification of specimens

Two swallow bug specimens (OV1 and OV2) were collected in May 2012 from a warehouse in Saskatoon. Specimens were identified morphologically as described in section 2.1.1. Diagnostically useful features included rostrum length, mesothoracic coxae separation, antennal segment length, as well as pronotal anterior margin and hairs (Figures 2.2 and Figure 2.6) (Usinger 1966). The DNA of each swallow bug was extracted and analyzed as described in section 2.1.2 (Table 2.10). The identity of the OV1 and OV2 was confirmed by performing PCR on their extracted genomic DNA as described in section 2.1.4.

TABLE 2.10
Spectrophotometric analysis of *Oeciacus vicarius* DNA Templates

Sample Code	260/280 Ratio[a]	Concentration (ng/µl)[b]
OV1	2.02	6.6
OV2	2.1	29.9

a: 260/280 is the ratio of absorbance at 260 and 280 nm, and is indicative of the purity of nucleic acids. For DNA, values falling below 1.8 can indicate contamination (NanoDrop®).
b: The concentration is the nanograms of DNA template per µl of AE buffer

Figure 2.6. Photographs of male *Oeciacus. vicarius* specimen showing the dorsal view (A) and ventral view (B).

2.3.2 Amplification and characterization of the Mt genome

Attempts were made to amplify the entire mitochondrial genome sequence of *Oeciacus vicarius* in two overlapping fragments as described in section 2.1.5. Despite troubleshooting each of these reactions, no DNA fragment was successfully amplified using either the Roche Expand Long Template PCR System or the Clontech Advantage® 2 PCR Enzyme System. PCR primers were designed for short range PCR as described in section 2.1.5. All of the primers that were successfully used to amplify mtDNA from *O. vicarius* are shown on Tables 2.11 and 2.12, and in Figure 2.6. Methods and troubleshooting for short range PCRs performed on the swallow bugs were conducted as described in section 2.1.5. All DNA analyses were performed as described in section 2.1.6.

TABLE 2.11
Forward PCR primers used to amplify the mtDNA of *Oeciacus vicarius*

Primer	Sequence	Position*
LEP-F	ATTCAACCAATCATAAAGATATNGG	1,316
Long F1	TGATAATTGGCGCACCTGATATAGCATTC	1,601
OV-NC-24F	ACTATAATTATCGCAGTCCC	2,270
Cimex-NC-4F	ACCACCCATATCATATAGTAGATTA	4,610
OV-NC-5F	ACAGGATTTCACGGTTTACATG	5,191
SER-F	GCTAACTATCTTTTAAAGCGG	6,021
OV-NC-6F	GCAGTAACAAGAAGTAGATGAATGG	7,219
OV-NC-21F	CAGCAGTAACAAGAAGTAGATG	7,220
Cimex-NC-37F	AACAGGAGCCTCAACATGAGC	8,648
Bbug-Mt-19F	ATACAGCTAATATTGAAATAGC	10,339
LRJ	TTACGCTGTTATCCCTAA	12,612
LongF2	AATAAATTTACCTGTCGCCCCAACAAAACT	12,703
Cimex-NC-41F	ACATTCAATAACAGAGGTATAC	14,029

***** The nucleotide position listed is the approximate position in relation to the mitochondrial DNA sequence of *Orius niger*.

TABLE 2.12
Reverse PCR primers used to amplify the mtDNA of *Oeciacus vicarius*

Primer	Sequence	Position*
OV-NC-22R	TTATAGATGGAGGGAGCAATC	1645
LEP-R	TAAACTTCTGGATGTCCAAAAAATCA	2,053
Bbug-Mt-24R	CTTGTTAATCCTCCAATTGTAATT	2,407
OV-NC-3R	AGAGTTATTGTTCCAATTGATCC	4,647
Cimex-NC-4R	ATCTTCCTCATCAATAAATGATA	5,352
OV-NC-SERR	TAAAGCGAAGCATTGAAGCTGC	6,083
OV-NC-6R	AAAGCGAAGCATTGAAGCTGC	6,084
Bbug-Mt-7R	GCTCAAATTCCTTTTTCTTC	7,312
OV-NC-33R	TCTTTGCCCCTTTTACTGGGTA	8,808
Cimex-NC-34R	ATCCTATCTATGATTTACAAGATC	9,584
Cimex-NC-m17R	GCAAATAGAAAATATCATTCTGG	10,983
mtBUG-16S-R	ATGGCCGCAGTACTTTAACTGTGCAAAGG	12,932
LRN	CGCCTGTTTATCAAAAACAT	13,009
Bbug-Mt-14R	TAAAAGAATTTGGCGGTAATT	14,057

***** The nucleotide position listed is the approximate position in relation to the mitochondrial DNA sequence of *Orius niger*.

3. RESULTS

3.1 Mitochondrial DNA sequence of *Cimex lectularius*

3.1.1. Mitogenome composition and organization

The arrangement of the mitochondrial genome (mitogenome) of *C. lectularius* based on aligned DNA sequence data was shown in Figure 3.1. A total of 14,927 bps of the mitogenome of *C. lectularius* was sequenced from over 30 overlapping mtDNA fragments (Figure 3.1). This was determined using 13,272 bps of specimen Cim17, 165 bps from specimen Cim18 and 1,490 bps from Cim20. The Cim18 sequence falls 19 bps upstream of the ND2 gene and the Cim20 sequence starts 258 bps upstream of the ND1 gene (Table 3.1). The mitogenome was AT rich (73%), with a nucleotide composition of A = 41%, T = 32%, G = 11%, and C = 16%.

Of the 13 putative protein coding genes (PCGs), 12 were sequenced in their entirety, while part of the 3' end of the CytB gene was not amplified (Table 3.1). Of the presumed 22 tRNA genes, 17 were detected in the sequences obtained (Table 3.1, Figure 3.1).

One region that could not be successfully amplified was an area spanning a portion of the control region and a region immediately upstream of the ND2 gene, that probably included three tRNA genes (Li et al. 2011; Li et al. 2012b; Yang et al. 2013; Kocher et al. 2015) (Figure 3.1). Though a ≈ 3kb fragment spanning part of the control region was successfully amplified, when it was sequenced the signal was too low, and the unreliable sequence was omitted from this thesis. A region between the ND1 and CytB genes of the mitogenome of *C. lectularius* could not be successfully amplified with traditional or long range PCR (Figure 3.1, Table 3.1). A $tRNA_{Leu}$ gene not part of the putative tRNA genes was detected overlapping and immediately upstream of the ND6 gene (Table 3.1). Additionally, the $tRNA_{Cys}$ expected

to be found upstream of the COI gene and between the $tRNA_{Trp}$ and $tRNA_{Tyr}$ genes was not present. There is a 51 bp intergenic space between the $tRNA_{Trp}$ and $tRNA_{Tyr}$ genes where the putative $tRNA_{Cys}$ gene normally is, but it was not found to code for a tRNA gene (Table 3.1). Last, a 258 bp region between ND1 and CytB that was successfully sequenced included an unusual repeat region resembling sequence from other parts of the mitogenome (Figure 3.2). With the exception of a single nucleotide, one 51 bp region of this repeat sequence is part of the ND1 gene. Another part of this repeating region is a sequence 102 bp in length that occurs immediately upstream and adjacent to the stop codon for the ND1 gene. With the exception of a single base pair difference, this 102bp sequence occurs in an intergenic space between the ND4l and ND6 genes, upstream of the $tRNA_{Leu}$ gene and overlapping it by 2 bp. All of the PCGs had either ATA or ATG as their start codon, and either TAA or TAG as their stop codon (Table 3.1).

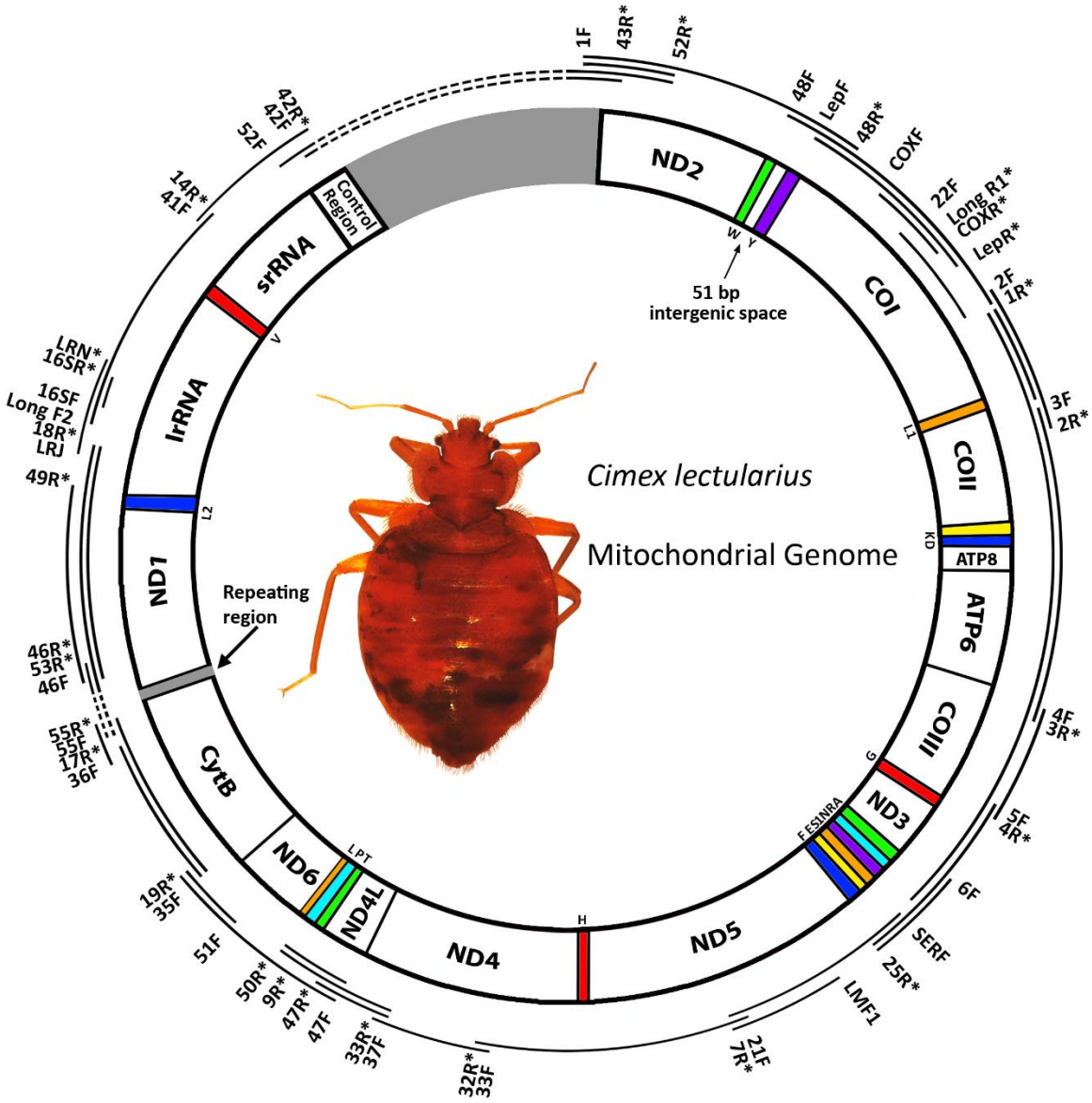
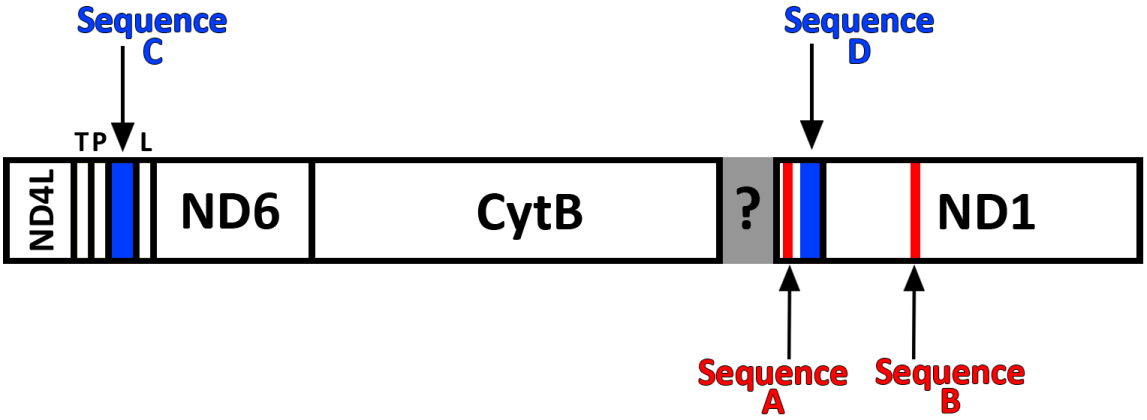


Figure 3.1. Mitochondrial map of *C. lectularius.* The position and length of the PCR fragments from which the DNA sequence was obtained are indicated by curved black lines in between forward and reverse primers. Reverse primers are signified with an asterisk. The tRNA genes are shown as colour blocked and labelled according to IUPAC-IUB single-letter amino acid codes. Partial fragments are indicated by dotted lines. The two regions of the mitogenome that have not been sequenced are shown in grey.




```
                          1-------------------------------------------------51
A - Upstream of ND1       TAATTCAGATTCTCCCTCAGCAAAATCGAAAGGAGAGCGGTTAGTTTCTGC
B - Within ND1            TAATTCAGATTCTCCCTCAGCAAAATCGAAAGGAGAGCGGTTGGTTTCTGC

                          1-------------------------------------------------52
C - B/T ND4L-ND6          TAAACAAAATAAATAAAGAGAAATATAATTTCCCCCCATAGGAGGGAAAAAA
D - Within ND1            TAAACAAAATAAATAAAGAGAAATATAATTTCCCCCCATAAGAGGGAAAAAA
                          53-----------------------------------------------102
C - B/T ND4L-ND6          TGGGGGGGATGTGAAATATATAAGTAAACAACAGCTTGTTTAGCTGCTAA
D - Within ND1            TGGGGGGGATGTGAAATATATAAGTAAACAACAGCTTGTTTAGCTGCTAA
```

Figure 3.2. Sequence A is 183 bp upstream of the stop codon for the ND1 PCG, sequence B is within the ND1 PCG, sequence C is between ND4l and ND6 PCGs, and sequence D is immediately upstream of ND1 stop codon. The tRNAs genes are labelled according to IUPAC-IUB single-letter amino acid codes. Differences in sequence have been highlighted in pink.

TABLE 3.1
Mitogenome annotations for *Cimex lectularius*

Gene	Length (base pairs)	Strand	Anticodon	Start Codon	Stop Codon
ND2	999	+		ATA	TAA
$tRNA_{Trp}$	69	+	TCA		
Intergenic space[a]	51				
$tRNA_{Tyr}$	65	-	GTA		
COI	1,539	+		ATA	TAA
$tRNA_{leu(UUR)}$	67	+	TAA		
COII	699	+		ATA	TAA
$tRNA_{Lys}$	75	+	CTT		
$tRNA_{Asp}$	63	+	GTC		
ATP8	144	+		ATA	TAG
ATP6	708	+		ATG	TAG
COIII	819	+		ATG	TAG
$tRNA_{Gly}$	63	+	TCC		
ND3	357	+		ATA	TAG
$tRNA_{Ala}$	64	+	TGC		
$tRNA_{Arg}$	67	+	TCG		
$tRNA_{Asn}$	67	+	GTT		
$tRNA_{Ser(AGN)}$	70	+	GCT		
$tRNA_{Glu}$	65	+	TTC		
$tRNA_{Phe}$	68	-	GAA		
ND5	1,677	-		ATA	TAA
$tRNA_{His}$	66	-	GTG		
ND4	1,326	-		ATG	TAA
ND4L	375	-		ATG	TAA
$tRNA_{Thr}$	65	+	TGT		
$tRNA_{Pro}$	69	-	TGG		
$tRNA_{Leu}$[b]	56	+	TAA		
ND6	522	+		ATA	TAA
CytB[c]	>1,212	+		ATA	?
ND1	930	-		ATA	TAA
$tRNA_{Leu(CUN)}$	67	-	TAG		
lrRNA	1,255	-			
$tRNA_{Val}$	70	-	TAC		
srRNA	784	-			

a: tRNACys gene is present in the ancestral mitochondrial genome in this spot
b: No tRNA present at this spot in ancestral insect mitogenome
c: Missing 3' end of this gene

3.1.2. Protein coding genes

The total length of all of the PCGs that were sequenced was 11,277, accounting for 75.5% of the total sequence composite for *C. lectularius* and coding for 3,758 amino acids (Table 3.1, Table 3.2). Each of the 13 PCGs had an ATR start codon, 9 being ATA (ND2, COI, COII, ATP8, ND3, ND5, ND6, CytB, and ND1) and 4 being ATG (ATP6, COIII, ND4, and ND4l) (Table 3.1). Four PCGs (ATP8, ATP6, COIII, and ND3) had a TAG stop codon, while the remainder had TAA. The stop codon for the CytB gene could not be obtained as it was within a region of the mitogenome that was not amplified by PCR (Table 3.1). The AT content of all PCGs is 72.2%.

TABLE 3.2
Amino acid composition of mitochondrial protein coding genes of *Cimex lectularius*

Amino Acid[a]	N[b]	Amino Acid	N	Amino Acid	N
F	346	T	185	D	67
L	530	A	144	E	83
I	347	Y	175	C	59
M	320	H	71	W	94
V	194	Q	59	R	53
S	378	N	189	G	206
P	134	K	124		

a: Amino acids are listed by their IUPAC-IUB single letter codes
b: Does not include stop codons

3.1.3. Ribosomal RNA and transfer RNA genes

The large ribosomal RNA (lrRNA) gene was 1,255 bps in length and had an AT content of 74.6%. This gene was positioned between the $tRNA_{leu(CUN)}$ and $tRNA_{Val}$ genes (Table 3.1). The small ribosomal RNA (srRNA) gene was 786 bps in length and had an AT content of 72.6%. The srRNA gene was positioned between the $tRNA_{Val}$ gene and the control region (Table 3.1). Both the rRNA genes were encoded by the N-strand. The total length of all the tRNA genes was 1,196 bps; the length of each tRNA gene ranged from 56 to 75 bps (Table 3.1). An intergenic space 51 bps in length lies between the $tRNA_{Trp}$ and $tRNA_{Tyr}$ genes (Table 3.1). No sequence was obtained of the region containing the $tRNA_{Ile}$, $tRNA_{Gln}$, $tRNA_{Met}$ and $tRNA_{Ser(TCN)}$ genes. A third putative $tRNA_{Leu}$ was detected between the $tRNA_{Pro}$ and ND6 genes (see Table 3.1). All of the secondary structures of the detected tRNAs displayed clover leaf shapes, with the exception of the unusual $tRNA_{Leu}$ that displayed a reduced dihydrouridine (DHU) arm and the $tRNA_{Ser(AGN)}$ gene, which has a loop DHU arm (Figure 3.3).

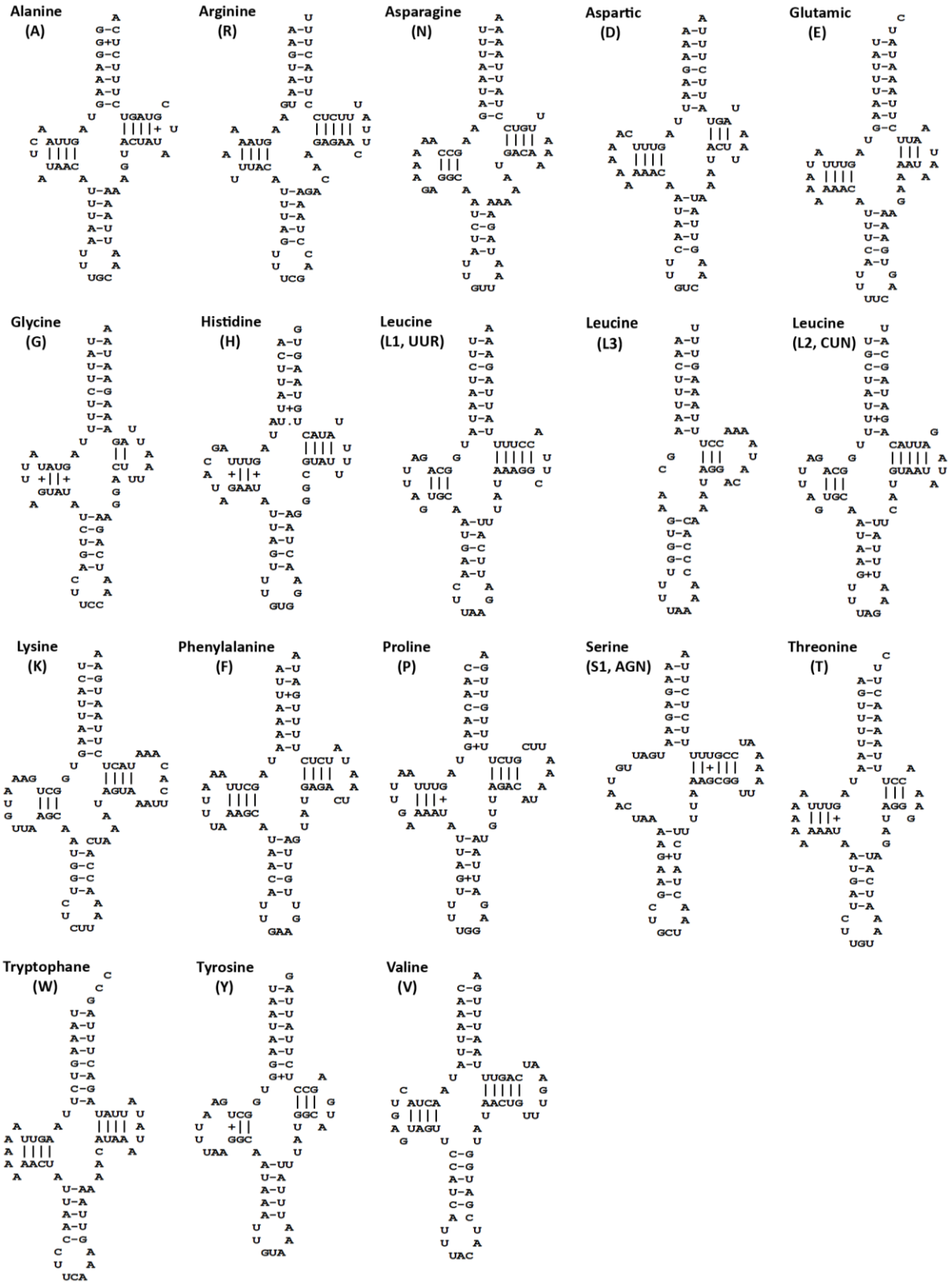


Figure 3.3. Inferred secondary structure of 18 of the putative 22 tRNAs genes of the *Cimex lectularius* mitogenome. Bonds drawn as lines show Watson-Crick base pairing and bonds drawn as a + indicate G-U base pairing.

3.1.4 Gene regions with intraspecific variation

Sequence variation could be examined among individuals for 4 genes; the ND2, COI, ND6 and ND5 (Table 3.3). There were 7 differences in a 271 bp nucleotide sequence of the ND2 gene between individuals Cim17 and Cim6. The ratio of transition to transversion point mutations in this sequence was 6:1. There were 3 purine transitions, 3 pyrimidine transitions, and 1 transversion. The mutational change at alignment position 11 corresponded to a change in the amino acid from a leucine in Cim17 to a methionine in Cim6.

There were 11 differences in a 552 bp nucleotide sequence of the COI gene between Cim17 and Cim6. The ratio of transition to transversion substitutions in this sequence was 10:1; 4 were purine changes, 6 were pyrimidine changes, and 1 was a transversion. The mutational change at position 1 corresponds to a change in amino acid from a valine in Cim17 to a leucine in Cim6 and the transition at position 187 corresponds to a change in amino acid from an adenine in Cim17 to a thymine in Cim6 (Table 3.3).

Cim17 and Cim6 also show variation in the ND6 gene, having 5 variable positions in a sequence 524 bps in length (ND6). The ratio of transition to transversion substitutions in this sequence was 4:1; 1 was a purine change, 3 were pyrimidine changes, and 1 was a transversion (Table 3.3). The transversion nucleotide difference at position 428 corresponds to a change in amino acid from a leucine in Cim17 to a phenylalanine in Cim6 (Table 3.3).

Sequence from the ND5 gene compared between Cim17 and Cim21 shows 6 bp differences across a sequence 225 bps in length. All point mutations in this sequence were transitions, 4 being purine transitions and the other 2 being pyrimidine. The transition at position 144 resulted in an amino acid substitution from methionine in Cim17 to valine in Cim21. The

change at position 225 resulted in an amino acid substitution from valine in Cim17 to isoleucine in Cim21.

3.1.5. Phylogenetic relationships of *Cimex lectularius*

A BLAST search of the amino acid sequence data of *Cimex lectularius* revealed that it was most genetically similar to species in Infraorder Cimicomorpha, with 8 of 13 PCGs having the top match being from this taxon (Table 3.4). Phylogenetic analyses were conducted on an amino acid sequence composed of 3,668 characters. Figure 3.4 shows the NJ tree produced from the analysis of amino acid sequence data for *Cimex lectularius* and 22 species of heteropterans (data obtained from Genbank). A MP analysis of the sequence data produced two equally most parsimonious trees (not shown) with a length of 17,043, consistency index (CI) based on 1,229 informative characters of 0.53 and a retention index (RI) of 0.39. *Cimex lectularius* belonged to a clade that included representatives of the Cimicomorpha and 3 other Infraorders of heteropterans. Excluded from this clade were two species of the Miroidea. There was 100% bootstrap support in both the NJ and MP trees for *C. lectularius* being placed in this clade. However, there was strong bootstrap support (97%) in the NJ tree against *C. lectularius* forming a clade with other species of the Cimicoidea (i.e. *Or. niger* and *Or. sauteri*). There was also no support in the bootstrap analysis of the MP trees for superfamily Cimicoidea forming a monophyletic clade.

TABLE 3.3
Gene regions of *Cimex lectularius* showing intraspecific variation

Protein coding gene	Sample code	Nucleotide positions of base pair differences*										
ND2		1	11[a]	28	55	196	268	271				
	Cim17	A	T	C	A	T	T	G				
	Cim6	G	A	T	G	C	C	A				
COI		1[a]	24	42	168	187[a]	282	330	399	465	504	552
	Cim17	G	A	G	G	G	C	C	C	C	T	G
	Cim6	T	G	A	A	A	T	T	T	T	C	A
ND6		1	200	245	428[a]	524						
	Cim17	A	T	T	A	C						
	Cim6	G	C	C	C	T						
ND5		1	31	58	103	144[a]	225[a]					
	Cim17	G	A	G	A	T	T					
	Cim21	A	G	A	G	C	C					

* Position 1 was determined by the first base pair difference observed between the two specimens in each gene.
a: Positions with point mutations causing amino acid substitutions

TABLE 3.4
BLAST results of the amino acid sequence of all 13 protein coding genes for *Cimex lectularius*

Protein Coding Gene	Species with Top Match	Infraorder	Superfamily	Query Coverage	Identity
ND2	*Peirates lepturoides*	Cimicomorpha	Reduvoidea	93%	46%
COI	*Gorpis humeralis*	Cimicomorpha	Naboidea	99%	86%
COII	*Sigara septemlineata*	Nepomorpha	Corixoidea	96%	71%
ATP8	*Halyomorpha halys*	Pentatomorpha	Pentatomoidea	100%	47%
ATP6	*Peirates fulvescens*	Cimicomorpha	Reduvoidea	100%	56%
COIII	*Physopelta gutta*	Pentatomorpha	Pyrrhocoroidea	84%	69%
ND3	*Apolygus lucorum*	Cimicomorpha	Miroidea	99%	66%
ND5	*Orius sauteri*	Cimicomorpha	Cimicoidea	98%	55%
ND4	*Enithares tibialis*	Nepomorpha	Notonectoidea	98%	64%
ND4L	*Himacerus nodipes*	Cimicomorpha	Naboidea	66%	59%
ND6	*Nerthra indica*	Nepomorpha	Ochteroidea	85%	49%
CytB	*Orius niger*	Cimicomorpha	Cimicoidea	99%	73%
ND1	*Peirates turpis*	Cimicomorpha	Reduvoidea	99%	71%

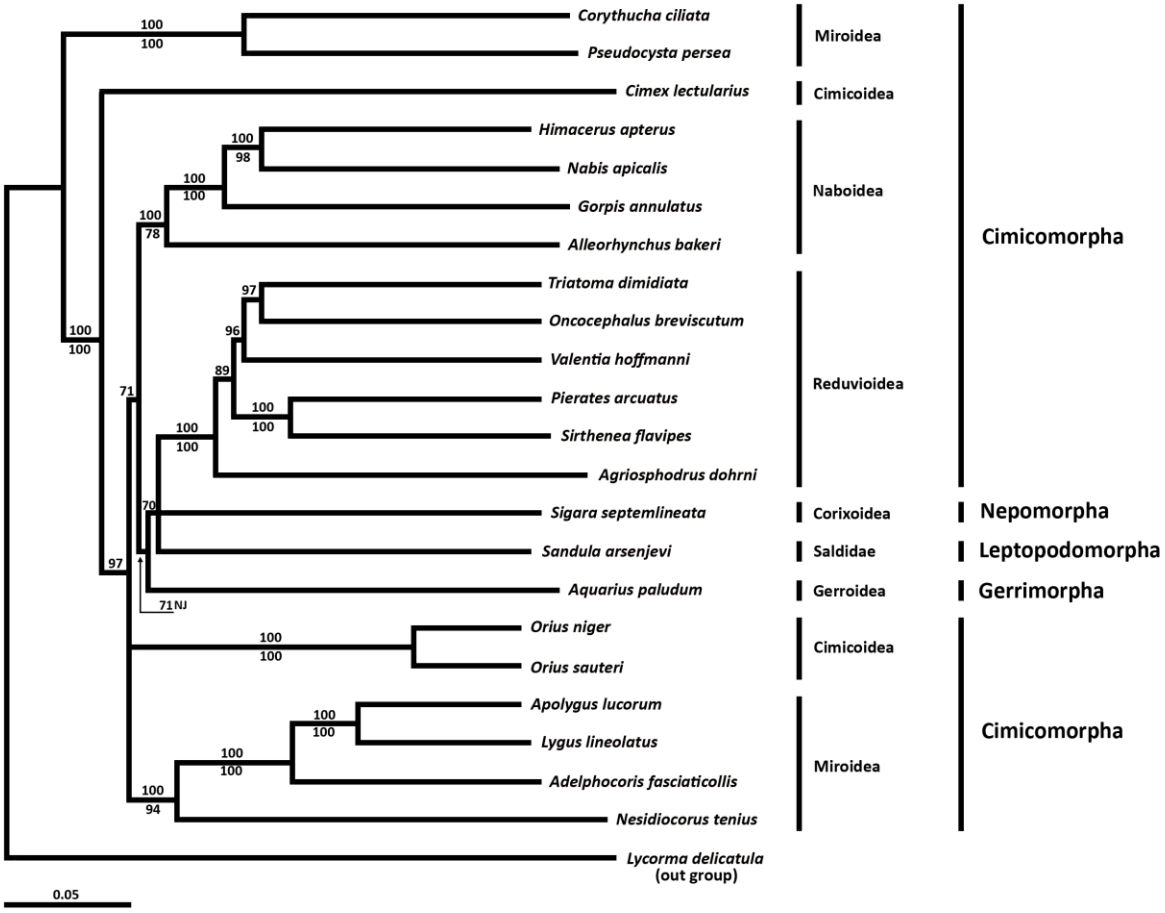


Figure 3.4. Phylogenetic relationship of *Cimex lectularius* to other heteropterans inferred from a neighbor joining analysis of 3,668 characters (amino acids) from 13 protein coding genes. Values above and below the branches are bootstrap values for the NJ and MP analyses, respectively.

3.2 Mitochondrial DNA sequence of *Cimex adjunctus*

3.2.1. Mitogenome composition and organization

The arrangement of the mitogenome of *C. adjunctus* based on aligned DNA sequence data is shown in Figure 3.5. A total of 4,898 bps of the mitogenome of *C. adjunctus* were sequenced from over 6 mtDNA fragments (Figure 3.5). This was determined using 2,477 bps of specimen Ca1, and 2,421 bps from specimen Bat1. The mitogenome was AT rich (68%), with a nucleotide composition of A = 36%, T = 32%, G = 15%, and C = 17%. Of the 13 putative protein coding genes (PCGs), 4 were partially sequenced (Table 3.5). Of the presumed 22 tRNAs, 4 were detected in the sequence obtained (Table 3.5 and Figure 3.5).

3.2.2. Protein coding genes

The total length of all of the PCGs that were sequenced was 3,201 bps, accounting for 65.4% of the total sequence composite for *C. adjunctus* and coding for 1,065 amino acids (Tables 3.5 and 3.6). The COI gene had ATG as a start codon, and the codons of the other PCGs did not fall in the sequence composite. The ND2 gene had TAA as a stop codon and was the only PCG to have its stop codon recognized in the sequence composite. The AT content of all PCGs is 79.5%.

3.2.3. Ribosomal RNA and transfer RNAs genes

A portion of the lrRNA gene 413 bps in length was sequenced and possessed an AT content of 72.6% (Table 3.5). No sequence was obtained of the regions containing the rest of the lrRNA gene or the snRNA gene. The total length of all the tRNA genes was 185 bps, having an AT content of 76.8%; the length of each tRNA gene ranged from 61 to 69 bps (Table 3.5). No sequence was obtained of the region containing the other tRNA genes. All secondary structures of the three detected tRNAs displayed clover leaf shapes (Figure 3.6).

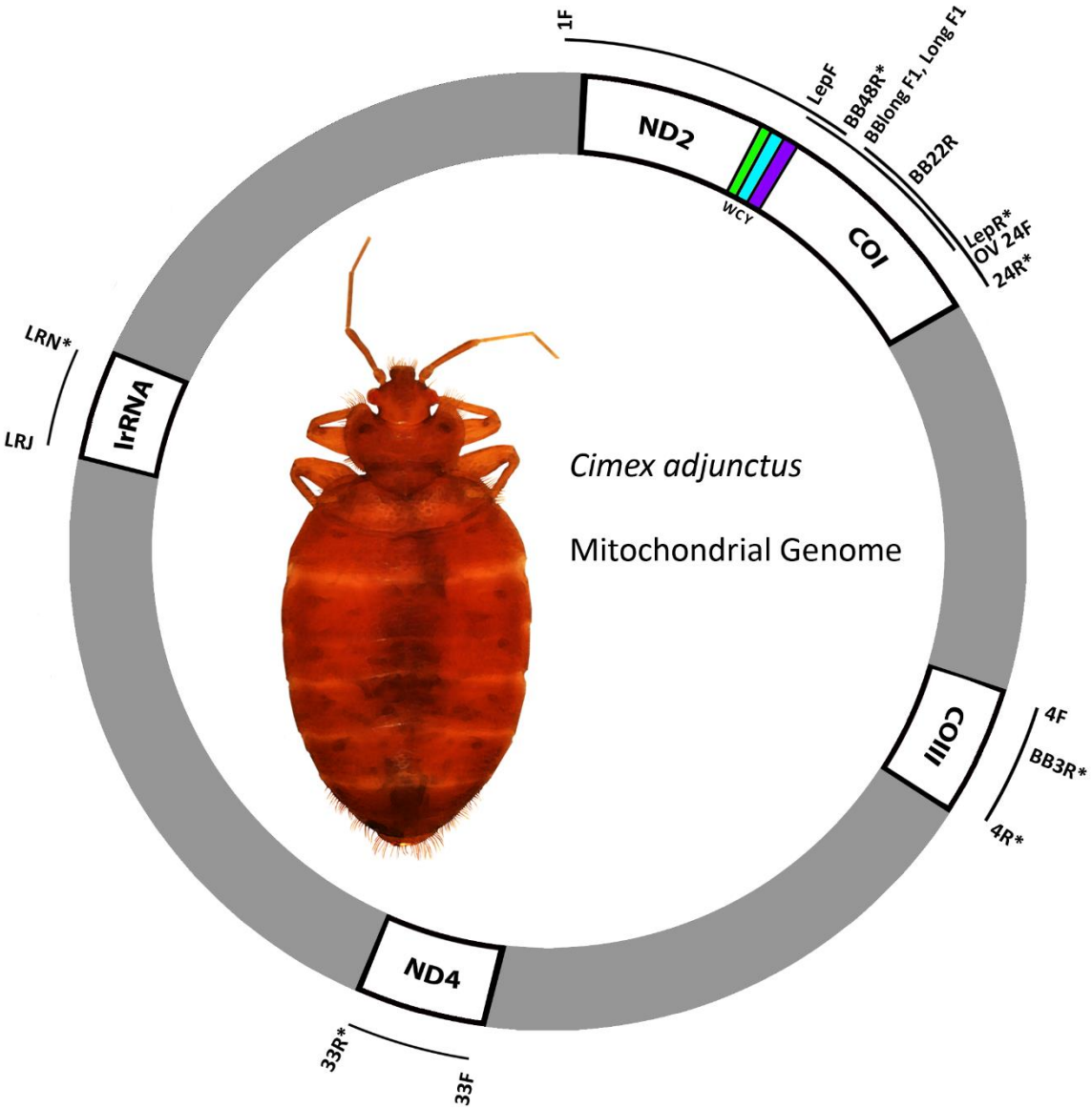


Figure 3.5. Mitochondrial map of *Cimex adjunctus.* The position and length of the PCR fragments from which the DNA sequence was obtained are indicated by curved black lines in between forward and reverse primers. Reverse primers are signified with an asterisk and tRNA genes were labelled according to the IUPAC-IUB single-letter amino acid codes. Regions of the mitogenome that have not been sequenced are shown in grey.

TABLE 3.5
Mitogenome annotations for *Cimex adjunctus*

Gene	Length (base pairs)	Strand	Anticodon	Start Codon	Stop Codon
ND2	>960	+		?	TAA
$tRNA_{Trp}$	69	+	TCA		
$tRNA_{Cys}$	61	-	GCA		
$tRNA_{Tyr}$	64	-	GTA		
COI	>921	+		ATG	?
COIII	>612	+		?	?
ND4	>708	-		?	?
lrRNA	>413	-			

TABLE 3.6
Amino acid composition of mitochondrial protein coding genes of *Cimex adjunctus*

Amino Acid[a]	N[b]	Amino Acid	N	Amino Acid	N
F	80	T	59	D	18
L	140	A	52	E	22
I	105	Y	31	C	13
M	103	H	26	W	30
V	51	Q	15	R	15
S	104	N	45	G	74
P	44	K	34	X[c]	4

a: Amino acids are listed by their IUPAC-IUB single letter codes
b: Does not include stop codons
c: Four of the amino acids were unknown to due unreliable sequence at these spots

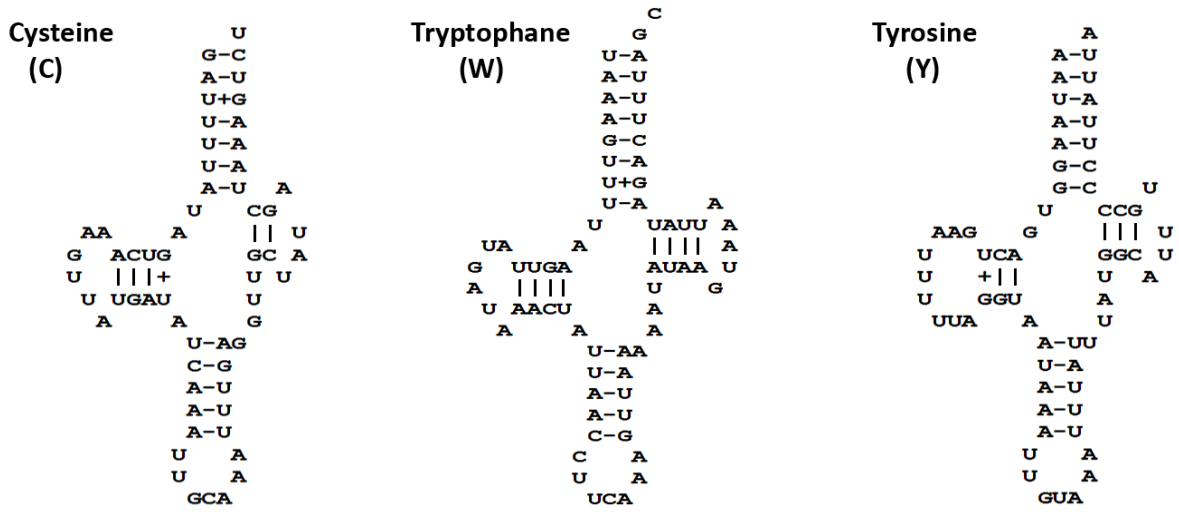


Figure 3.6. Inferred secondary structure of 3 of the putative 22 tRNA genes of the *Cimex adjunctus* mitogenome. Bonds drawn as lines show Watson-Crick base pairing and bonds drawn as a + indicate G-U base pairing.

3.2.4. Phylogenetic relationships of *Cimex adjunctus*

Phylogenetic analyses were conducted on an amino acid sequence composed of 1,057 characters. Figure 3.7 shows the NJ tree provided for the analysis of amino acid sequence data for *C. adjunctus*, *C. lectularius* and 22 species of heteropterans (data obtained from Genbank). A MP analysis of the sequence data produced three equally most parsimonious trees (not shown) with a length of 4,671, a consistency index (CI) based on 589 informative characters, of 0.54 and an RI of 0.41 (Figure 3.7). There was 100% bootstrap support in both the NJ and MP trees for *C. adjunctus* and *C. lectularius* belonging to the same clade, to the exclusion of all other species. *Cimex adjunctus* and *C. lectularius* belonged to a clade that included representatives of the Cimicomorpha and 3 other Infraorders of heteropterans. Excluded from this clade were two species of the Miroidea. There was 92% bootstrap support in the NJ and 72% in the MP tree for *C. adjunctus* and *C. lectularius* being placed in this clade. However, there was bootstrap support (79%) in the NJ tree against *C. adjunctus* and *C. lectularius* forming a clade with other species of the Cimicoidea (i.e. *Or. niger* and *Or. sauteri*). There was also no support for Cimicoidea forming a monophyletic clade in the MP trees.

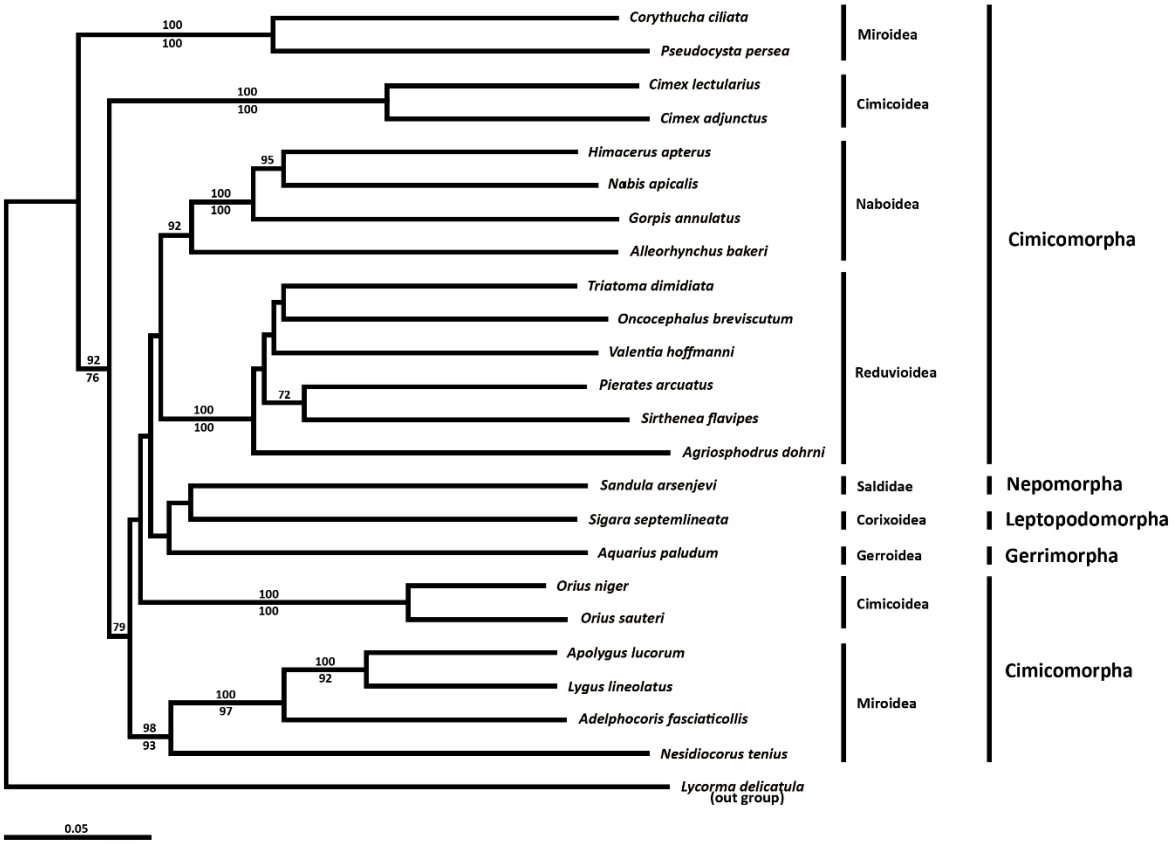


Figure 3.7. Phylogenetic relationship of *C. lectularius* and *C. adjunctus* to other heteropterans inferred from a neighbor joining analysis of 1,057 characters (amino acids) from the ND2, COI, COIII, and ND4 mt protein coding genes. Values above and below the branches are bootstrap values for the NJ and MP analyses , respectively.

3.3 Mitochondrial DNA sequence of *Oeciacus vicarius*

3.3.1. Mitogenome composition and organization

The arrangement of the mitochondrial genome (mitogenome) of *O. vicarius* based on aligned DNA sequence data is shown in Figure 3.8. A total of 8,751 bps of the mitogenome of *O. vicarius* were sequenced from over 11 mtDNA fragments (Figure 3.9). All of this was determined using sequence data from specimen OV2. The mitogenome was AT rich (69%), with a nucleotide composition of A = 40%, T = 29%, G = 13.5%, and C = 17.5%. Of the 13 putative protein coding genes (PCGs), 4 were sequenced in their entirety, while 3 were partially sequenced (Table 3.7). Of the presumed 22 tRNAs, 4 were detected in the sequence obtained (Figure 3.8).

3.3.2. Protein Coding Genes

The total length of all of the PCGs that were sequenced was 6,582, accounting for 75.2% of the total sequence composite for *O. vicarius* and coding for 1,967 amino acids (Tables 3.7 and 3.8). The COIII, ND3 and ND5 genes each had ATA as a start codon, while the ND4 and ND4L genes used an ATG start codon (Table 3.7). The COIII, ND5 and ND4L gene use TAA as a stop codon, while the ND3 and ND4 genes use a TAG stop codon. Start and stop codons for the remaining PCGs were not detected in the sequence composite. The AT content of all sequenced PCGs is 69.5%.

3.3.3. Ribosomal RNA and transfer RNA genes

A portion of the lrRNA gene 906 bps in length was sequenced, possessing an AT content of 74% (Table 3.7). A portion of the srRNA gene 338 bps in length was sequenced, possessing an AT content of 66.3%. The total length of all the tRNA genes was 532 bps and an AT content of 75.9%; the length of each tRNA gene ranged from 63 to 70 bps. No sequence was

obtained of the regions presumed to contain the other tRNA genes. The $tRNA_{His}$ gene present in most other insect mitogenomes was not detected, but an intergenic space 55bps in length lies between the ND5 and ND4 genes. All of the secondary structures of the detected tRNAs displayed clover leaf shapes.

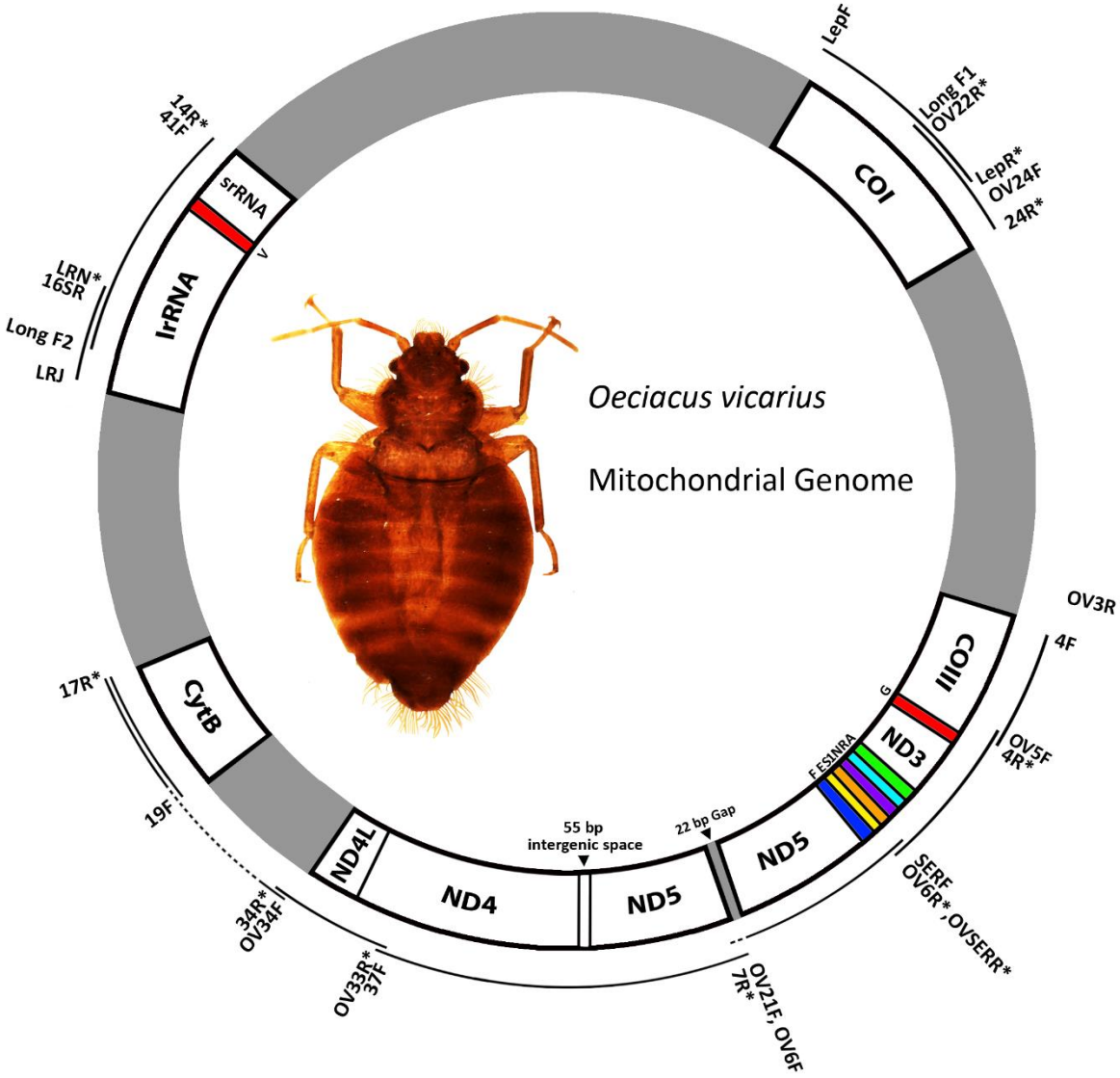


Figure 3.8. Mitochondrial map of *Oeciacus vicarius.* The position and length of the PCR fragments from which the DNA sequence was obtained are indicated by curved black lines in between forward and reverse primers. Reverse primers are signified with an asterisk and tRNA genes are labelled according to the IUPAC-IUB single-letter amino acid codes. Partial fragments are indicated by dotted lines, while regions of the mitogenome that have not been sequenced are shown in grey.

TABLE 3.7
Mitogenome annotations for *Oeciacus vicarius*

Gene	Length (base pairs)	Strand	Anticodon	Start Codon	Stop Codon
COI	>885	+		?	?
COIII	>780	+		?	?
$tRNA_{Gly}$	63	+	TCC		
ND3	366	+		ATA	TAG
$tRNA_{Ala}$	66	+	TGC		
$tRNA_{Arg}$	65	+	TCG		
$tRNA_{Asn}$	68	+	GTT		
$tRNA_{Ser(AGN)}$	70	+	GCT		
$tRNA_{Glu}$	65	+	TTC		
$tRNA_{Phe}$	66	-	GAA		
ND5	>1,656	-		ATA	TAA
ND4	1,335	-		ATG	TAG
ND4L	282	-		ATG	TAA
CytB	>609	+		?	?
lrRNA	>906	-			
$tRNA_{Val}$	70	-	TAC		
srRNA	>338	-			

TABLE 3.8
Amino acid composition of mitochondrial protein coding genes of *Oeciacus vicarius*

Amino Acid[a]	N[b]	Amino Acid	N	Amino Acid	N
F	210	T	91	D	40
L	289	A	75	E	40
I	170	Y	89	C	37
M	158	H	44	W	50
V	125	Q	23	R	29
S	186	N	71	G	135
P	68	K	37		

a: Amino acids are listed by their IUPAC-IUB single letter codes
b: Does not include stop codons

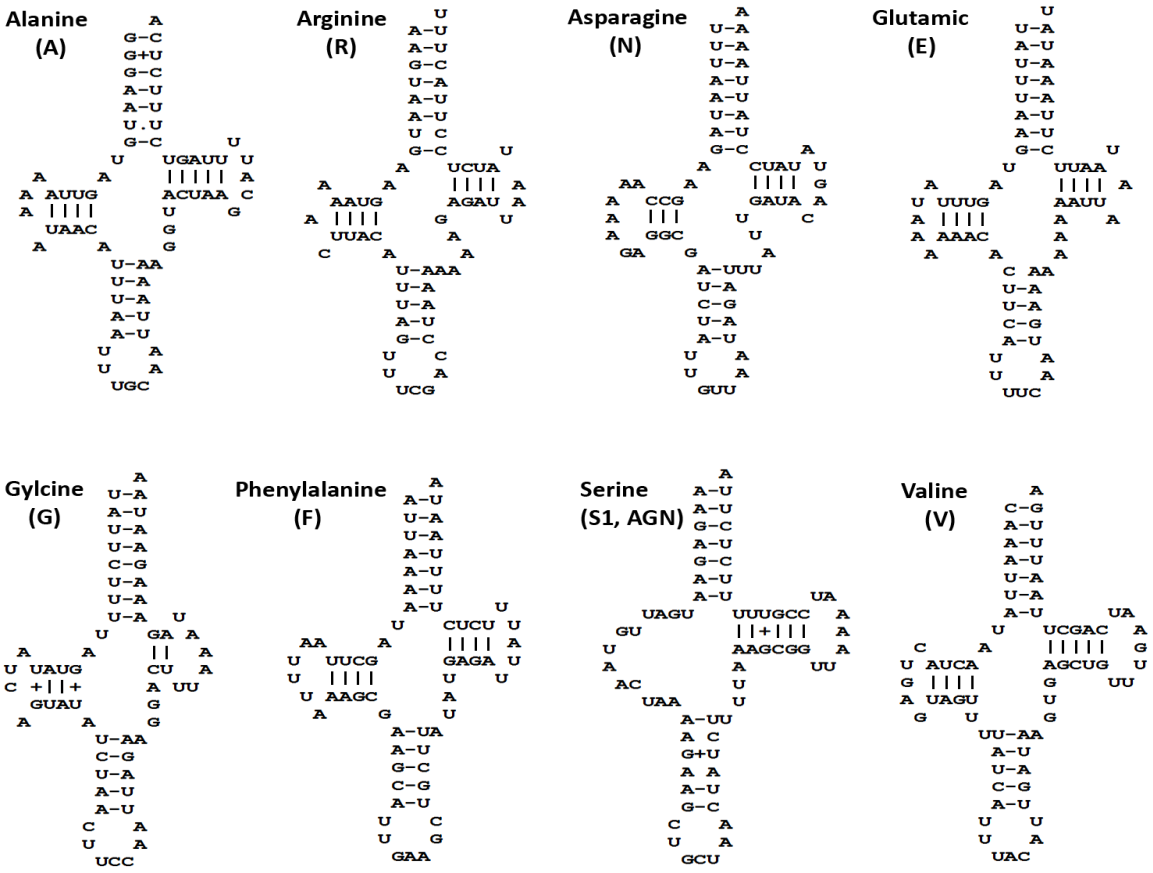


Figure 3.9. Inferred secondary structure of 8 of the putative 22 tRNA genes of the *Oeciacus vicarius* mitogenome. Bonds drawn as lines show Watson-Crick base pairing and bonds drawn as a + indicate G-U base pairing.

3.3.4. Phylogenetic relationships of *Oeciacus vicarius*

Phylogenetic analyses were conducted on an amino acid sequence composed of 1,829 characters. Figure 3.10 shows the NJ tree provided for the analysis of amino acid sequence data for *O. vicarius*, *C. lectularius* and 22 species of heteropterans (data obtained from Genbank). A MP analysis of the sequence data produced three equally most parsimonious trees (not shown) with a length of 7,995, a CI, based on 1,019 informative characters, of 0.54 and an RI of 0.42 (Figure 3.10). There was 100% bootstrap support in both the NJ and MP trees for *O. vicarius* and *C. lectularius* belonging to the same clade, to the exclusion of all other species. *Oeciacus vicarius* and *C. lectularius* belonged to a clade that included representatives of the Cimicomorpha and 3 other Infraorders of heteropterans. Excluded from this clade were two species of the Miroidea. There was 99% bootstrap support in the NJ and 83% in the MP tree for *O. vicarius* and *C. lectularius* being placed in this clade. However, there was bootstrap support (80%) in the NJ tree against *O. vicarius* and *C. lectularius* forming a clade with other species of the Cimicoidea (i.e. *Or. niger* and *Or. sauteri*). There was also no support for Cimicoidea forming a monophyletic clade in the MP trees.

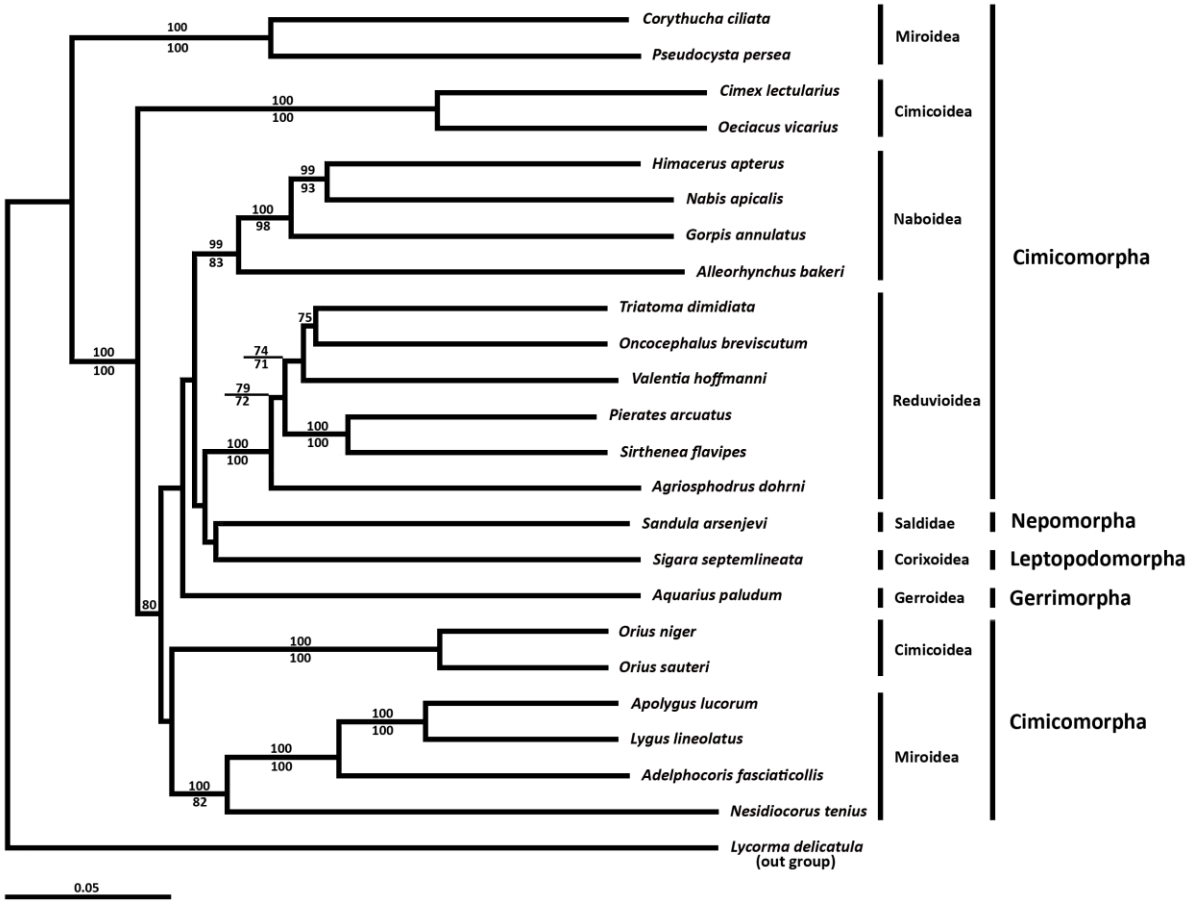


Figure 3.10. Phylogenetic relationship of *Cimex lectularius* to *Oeciacus vicarius* and other heteropterans inferred from a neighbor joining analysis of 1,829 characters (amino acids) from the COI, COII, ND3, ND4, ND5, ND4L, and CytB mt protein coding genes. Values above and below the branches are bootstrap values for the NJ and MP analyses , respectively.

3.4. Phylogenetic relationships of *C. lectularius, C. adjunctus* and *O. vicarius*

The nucleic and amino acid sequences of *C. lectularius, C. adjunctus, O. vicarius and Or. niger* were aligned over 2,595 and 727 characters, respectively. The number of differences in nucleotide sequence between these species ranged from 466 to 641, while the number of differences in amino acid sequence ranged from 38 to 101 (Table 3.9).

Phylogenetic analyses were conducted on a nucleotide sequence composed of 2,595 characters. Figure 3.11 shows the NJ tree provided for the analysis of nucleotide sequence data for *C. lectularius, C. adjunctus,*and *O. vicarius*. A MP analysis of the sequence data produced 1 most parsimonious tree (not shown) with a length of 1,323, a CI, based on 176 informative characters, of 0.63 and an RI of 0.42. *Cimex lectularius* and *O. vicarius* belonged to a clade that excluded *C. adjunctus*. There was strong bootstrap support for both the NJ and MP trees (98% and 91%, respectively) for *C. lectularius* and *O. vicarius* being placed in this clade. There was also no support for *C. adjunctus* belonging in the same clade as the other two cimicids.

Phylogenetic analyses were conducted on an amino acid sequence composed of 727 characters. Figure 3.12 shows the NJ tree provided for the analysis of amino acid sequence data for *C. lectularius, C. adjunctus,* and *O. vicarius*. A MP analysis of the sequence data produced 1 most parsimonious tree (not shown) with a length of 333, a CI, based on 18 informative characters, of 0.62 and a retention index (RI) of 0.39. *Cimex lectularius* and *O. vicarius* belonged to a clade that excluded *C. adjunctus*. There was strong bootstrap support for both the NJ tree (94%) for *C. lectularius* and *O. vicarius* being placed in this clade. There was also no support for *C. adjunctus* belonging in the same clade as the other two cimicids.

TABLE 3.9
Number of nucleotide (lower diagonal) and amino acid (upper diagonal) differences in the DNA sequence between *Cimex lectularius, Cimex adjunctus, Oeciacus vicarius,* and *Orius niger*

	C. lectularius	*C. adjunctus*	*O. vicarius*	*Or. Niger*
C. lectularius	-	61	38	101
C. adjunctus	507	-	64	99
O. vicarius	466	512	-	94
Or. niger	641	633	631	-

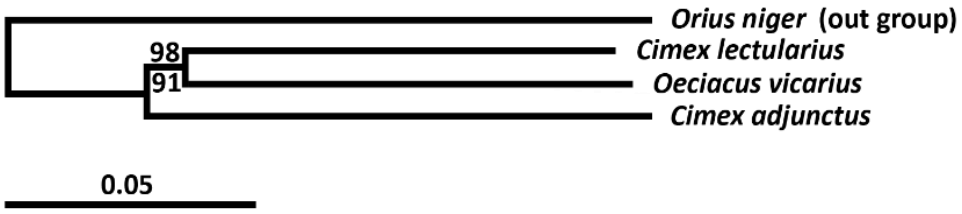


Figure 3.11. Phylogenetic relationship of *C. lectularius*, *C. adjunctus* and *O. vicarius* inferred neighbor joining analysis. The tree was inferred using 2,595 nucleotide characters across their respective mitogenomes. Only bootstrap values greater than 70% are listed above the branches, while maximum parsimony bootstrap values are listed as percentages below the branch.

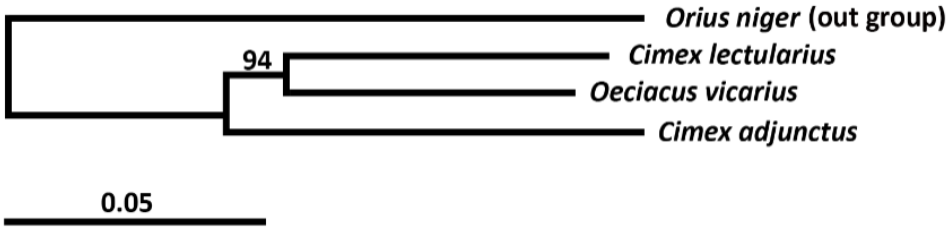


Figure 3.12. Phylogenetic relationship of *C. lectularius*, *C. adjunctus* and *O. vicarius* inferred by neighbor joining analysis. The tree was inferred using 727 amino acid characters across 4 PCGs. Only bootstrap values greater than 70% are listed above the branches.

4. DISCUSSION

4.1 Gene arrangements and composition

In the present study, the mitochondrial arrangements, gene directions, start and stop codons, and tRNA secondary structures for *C. lectularius*, *C. adjunctus* and *O. vicarius*, are mostly the same as the hypothesized ancestral insect mitogenome (Li et al. 2011; Li et al. 2012a; Li et al. 2012b; Cameron 2014b; Kocher et al. 2015). This is consistent with the literature, as mitochondrial genome changes in gene order or direction are rare within suborder Heteroptera (Cameron 2014b).

The mitochondrial genome of *C. lectularius* has all 13 protein coding genes, two rRNA genes, and 17 of the 22 expected tRNA genes (Cameron 2014a; Kocher et al. 2015). However, the $tRNA_{Ile}$, $tRNA_{Gln,}$ and $tRNA_{Met}$ genes downstream of the CR and upstream of the ND2 gene in the putative insect mitogenome (Cameron 2014a; Kocher et al. 2015) were not detected in the mitochondrial genome of *C. lectularius* due to a lack of sequence data in this region. The same explanation applies for the detection of the expected $tRNA_{Ser}$ gene between the ND1 and CytB genes (Cameron 2014a; Kocher et al. 2015).

A notable deviation from the putative insect mitochondrial genome in *C. lectularius* is the absence of the $tRNA_{Cys}$ between $tRNA_{Trp}$ and $tRNA_{Tyr}$; instead, there was a 51bp intergenic spacer. Another anomaly was the detection of a $tRNA_{Leu}$ gene not found in the hypothesized ancestral insect mitogenome, located upstream and overlapping the ND6 gene by 18 bp (Figure 3.1 and Table 3.1). The secondary structure inferred from this $tRNA_{Leu}$ was found to have a reduced DHU arm (Figure 3.4). Based on the putative insect genome it was expected that a $tRNA_{His}$ gene (Cameron 2014a; Kocher et al. 2015) would be found in *O. vicarius*, between the ND5 and ND4 genes, but a 55 bp intergenic spacer occupied this

region (Figure 3.6 and Table 3.5). Of the insect species to have had their mitochondrial genome sequenced, very few have lost mitochondrial tRNA genes, and most of these are lice species in the Order Phthiraptera (Cameron 2014b). Furthermore, the inference of tRNA genes based on secondary structure does not detect tRNA gene isotypes that do not conform to a cloverleaf structure (Beckenbach and Joy 2009; Cameron 2014b). With this information in mind, it is likely that the lack of detection of expected tRNA genes does not indicate their absence, and that the 51 and 55 bp intergenic spacers found in *C. lectularius* and *O. vicarius* are part of the putative $tRNA_{Cys}$ and $tRNA_{His}$, respectively. There is also evidence that mitochondrial tRNAs inferred from secondary structure, that deviate from the putative tRNAs, can be spurious, especially if they overlap with open reading frames of adjacent genes (Cameron 2014b). Given that the unexpected $tRNA_{Leu}$ detected in the mitogenome of *C. lectularius* has an 18 bp overlap with the ND6 gene, the hypothesized $tRNA_{Leu}$ gene is almost certainly specious (Cameron 2014b).

The region between the ND1 and CytB of *C. lectularius'* mitogenome contained an unsequenced gap, and two repeating sequences, which is usually abnormal in the Heteroptera (Cameron 2014a; Kocher et al. 2015). However, the mitogenome of the kissing bug, *Triatoma dimidiata*, was found to contain an unusual 314 bp tandem repeat intergenic spacer located between the ND1 and CytB genes (Dotson and Beard 2001) in the same spot (Figure 3.2). Nonetheless, the DNA sequences of this intergenic repeat region in these Cimicoidean species are quite different. When the 258 bp repeat sequence of *C. bclectularius* was compared to available sequence data on GenBank there were no matches, except for a partial match to the DNA sequence of the ND1 gene. However, when the sequence is entered into a BLAST of the EST database there were perfect matches to sequence data for *C.lectularius*,

suggesting that the sequence data produced from this study was not an artefact, but consistent with the findings of the mt sequence data of *C. lectularius* in the study by Moriyama et al. (2012). Moreover, other parts of the intergenic repeat region were also detected in sequence data from an incomplete shotgun assembly of the complete genome of *C. lectularius* (Qu et al. 2015). It corresponded to the sequence positions 13,503,659 to 13,504,390 of the complete genome of *C. lectularius* (Qu et al. 2015).

It has been suggested that the intergenic repeat region in *Triatoma dimidiata* may be a second origin of replication, similar to the role a 138 bp intergenic spacer plays in *Apis mellifera* (Crozier and Crozier 1993; Dotson and Beard, 2001). Perhaps the repeat region in *C. lectularius* also represents a second origin of replication; however, it is not as AT rich as expected if it was to function as such (Crozier and Crozier 1993; Dotson and Beard 2001). It is also possible that this sequence data is a nuclear copy of ex-mitochondrial DNA (Numt) (Dotson and Beard 2001). Numts can be as small as 200 bp, and if the primers used to amplify this region have binding sites on both the mitogenome and a Numt it may explain why PCRs of this sequence are so difficult (Dotson and Beard 2001). The significance of a 102 bp region occurring in both this repeat region and an intergenic space between the ND4l and ND6 genes needs to be explored further.

4.2 Failed PCRs

The gene regions that were not successfully sequenced in each of the three cimicids are due to failed PCRs. Failure to amplify targeted DNA sequence is usually attributed to various aspects of the PCR conditions, such as Mg^{++} concentration, buffer pH, primer design, DNA template quality, difficulty of target, and cycling conditions (Davies and Gray 2002; Jones 2002; Roux 2009). Cycling conditions and reagent concentrations were considered during

the trouble shooting of the PCRs (Chapter 2), so it is unlikely these factors contributed to the failed PCRs. Primers were designed to be specific to the target, and most of the primers used had sufficient GC content and were checked for self-annealing, cross dimer and hairpin formation and runs of 4 or more of the same base, making it unlikely that poor primer design was a cause for the failed PCRs (Davies and Gray 2002; Jones 2002; Roux 2009).

The repeated failures to obtain high yield PCR product using a long PCR protocol for two different kits, and a variety of thermal cycling conditions and reagent concentrations as described in Chapter 2, may be explained by poor template quality. Particularly long or difficult targets require high quality template, as the quantity of full length targets that are un-nicked decreases as the target length increases (Clontech, 2009). This can result in failed PCRs, or in truncated products due to depurination during thermal cycling (Clontech, 2009). The DNA extraction described in Chapter 2 resulted in 100 µl template elutions with concentrations ranging from 9 to 20 ng/µl. The long PCR protocol used in the present study recommends that 100 ng of DNA be used as the template, hence trouble shooting the conditions for DNA samples based on a single individual was costly on available template resources, limiting the number of attempts (Clontech, 2009). The 260/280 values expected from a solution containing pure DNA is 1.8 to 2 (NanoDrop®), yet most of the tested templates were greater than 2. Values that fall below 1.8 suggest there is protein contamination, making contamination of template unlikely (NanoDrop®). The abnormally high values could be due to an overall AT rich skew, since adenine has a much higher 260/280 than the other nucleic acids found in DNA (NanoDrop®). Alternatively, it could also signal RNA contamination, as uracil also has a very high 260/280 (NanoDrop®).

Other areas that were difficult to amplify with Biorad (Bio-Rad Laboratories, California, USA) or Fermentas (Fermentas, Massachusetts, USA) Taq polymerase, such as the repeating region, were successfully amplified with a high yield using Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific Inc. 2011). This enzyme has proof reading activity and PCRs performed with this enzyme consistently produced products of a higher yield, size and specificity than the latter two polymerases (Thermo Fisher Scientific Inc. 2011). Despite the success in amplifying 258 bp of this region, an unknown amount remained unsequenced due to PCR failure (Figures 3.1 and 3.2). There is also missing sequence in a partial shotgun assembly of *C. lectularius* DNA in the same spot as reported in the present study, suggesting it is a particularly difficult region to amplify (Qu et al. 2015).

4.3 Intraspecific variation in DNA sequences

Previous population genetic studies on *C. lectularius* have used sequence data from the ITS rDNA, the mt COI and 16S genes, and microsatellite loci (Szalanski et al. 2008; Booth et al. 2011; Vargo et al. 2011; Balvin et al. 2012; Davies et al. 2012; Fountain et al. 2014; Booth et al. 2015). However, inclusion of additional genetic markers may add to the reliability of the conclusions drawn from these population genetic studies. The results of the present study revealed that the regions of the mitogenome that differed in sequence between *C. lectularius* individuals from two different infestations in Saskatoon included the COI, ND2, ND5 and ND6 genes. Further study is needed to establish if one or more of these genes are useful in population genetics studies of *C. lectularius*. In addition to these genes, the DNA sequences of the entire mitochondrial genome of *C. lectularius* may be useful for population genetics studies (Ma et al. 2012; Cameron 2014a).

The number of transitional changes in DNA sequence among individual bed bugs was much greater than the number of transversions in each of the genes examined, which is consistent with findings of the mitogenomes of other insects (Yang and Yoder 1999; Keller et al. 2007). Furthermore, six of the 29 detected nucleotide differences in DNA sequences among C. lectularius individuals corresponded to changes in amino acid sequence. Though transitions are less likely to cause amino acid substitutions due to "wobble" (Yang and Yoder 1999; Keller et al. 2007), three of the six nucleotide differences were transitional changes. It is unlikely these substitutions are deleterious (Ng and Henikoff, 2006).

4.4 Phylogenetic relationships of bed bugs and their relatives

The genetic similarity of *C. lectularius* with other species in infraorder Cimicomorpha is consistent with the established taxonomic and phylogenetic relationships (Usinger 1966; Schuh et al. 2009). Phylogenetic analysis using amino acid sequence data have shown that the Cimicomorpha does not represent a monophyletic assemblage. This is consistent with the literature using complete mitochondrial genomes, but differs from analyses focused on morphological characters and limited nuclear and mitochondrial sequence data (Schuh et al. 2009; Wierauch and Schuh 2010; Kocher et al. 2015).

Phylogenetic analyses have also shown Miroidea as not representing a monophyletic assemblage, which is in agreement with other studies using complete mitochondrial genomes, but is at odds with studies using limited molecular and morphological data (Schuh et al. 2009; Wierauch and Schuh 2010; Kocher et al. 2015). Previous phylogenetic studies using complete mitochondrial DNA were limited to one or two species within the genus *Orius* to represent superfamily Cimicoidea, leaving the relationships within this superfamily unresolved (Schuh et al. 2009; Wierauch and Schuh 2010; Kocher et al. 2015). Despite

evidence from previous studies, based on morphological characters and limited molecular data, which indicated that the Cimicoidea was monophyletic (Schuh et al. 2009; Wierauch and Schuh 2010; Kocher et al. 2015), there was strong support in the present study that Cimicoidea is not a monophyletic clade. Furthermore, there was no support in the present study to conclude Cimicoidea is monophyletic. In contrast, the Naboidea and Reduvoidea were found to be monophyletic, which is consistent with previous research (Schuh et al. 2009; Wierauch and Schuh 2010; Kocher et al. 2015).

There was complete support for *C. lectularius*, *C. adjunctus* and *O. vicarius* being in the same clade, to the exclusion of the genus *Orius*. Consistent with this is the greater number of nucleotide differences between each of the cimicids and *Orius niger*. The phylogenetic analysis of the three cimicids with each other using both nucleic acid and amino acid characters showed strong support for *C. lectularius* as being the sister taxon to *O. vicarius*, and not to *C. adjunctus*, despite their taxonomic classification (i.e. in the same genus; Usinger 1966). The number of nucleic and amino acid differences detected between *C. adjunctus* and *C. lectularius* was also greater than between *O. vicarius* and *C. lectularius*, further supporting the closer association of *C. lectularius* with *O. vicarius*. *Cimex* and *Oeciacus* are traditionally considered sister taxa (Balvin et al. 2015). Given that the current phylogenetic and taxonomic relationships within family Cimicidae are based on host relationships and morphology, it is possible that the inclusion of molecular data could cause a restructuring of the systematic relationships (Usinger 1966; Balvin et al. 2015). At the time of writing this thesis, a paper was published that examined the phylogenetic relationship of *O. vicarius* to other cimicids using the 16S, 18S, COI and EF1α genes, and morphological data (Balvin et al. 2015). It was found that the genus *Cimex* was not a monophyletic assemblage

as *O. vicarius* and *O. hirundinis* were placed internally within this clade (Balvin et al. 2015). However, in the study by Balvin et al. (2015), *C. adjunctus* was placed closer to *C. lectularius* than to *O. vicarius.* This suggests that the choice of genetic markers can have an important impact on phylogenetic inference.

4.5 Conclusions and future research

As far as I am aware, the present study was the first attempt to determine the DNA sequence of the complete mitochondrial genome of an individual *C. lectularius.* Furthermore, it was the first study to determine the sequence of a nearly complete mitogenome of a representative of a species in the family Cimicidae (Balvin et al. 2015; Kocher et al. 2015). The unexpected repeat region and unsequenced gap between the ND1 and CytB genes needs further investigation, as it is unknown if it is an origin of replication, a Numt, or something else entirely. Attempts to amplify this region using PCR and DNA from isolated mitochondria (Chapter 2) could rule out the possibility of this region being nuclear DNA. Sequences of *O. vicarius, C. adjunctus* and other cimicids should also be obtained from this region of their mitogenomes to determine if this repeat region is species-specific, or exists throughout family Cimicidae, as this could be used to infer evolutionary relationships within this important family. High fidelity polymerase and quality DNA template have been shown to be vital in amplifying long or difficult targets. It is recommended that future research use the Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific Inc. 2011). The present study has designed many primers to effectively amplify the mt DNA of *C. lectularius, C. adjunctus* and *O. vicarius,* and many of these primers should also work on closely related species. This research has also found molecular markers that are potentially useful for population genetics studies, located in regions other than the mt 16S and COI genes used

previously. At an infraorder and superfamily level, the phylogenetic analyses are inconsistent with established phylogenies inferred from morphological characters and limited nuclear and mitochondrial sequence data (Schuh et al. 2009; Wierauch and Schuh 2010; Kocher et al. 2015). On the other hand, the finding that Cimicomorpha is not monophyletic is congruent with other studies using partial or complete mitochondrial genome sequence data (Li et al. 2012a; Li et al. 2012b; Kocher et al. 2015). It has been suggested that phylogenetic studies relying exclusively on mt DNA have limitations, so future research should include also data from nuclear genes in order to resolve the open ended nature of these inconsistent studies (Cameron 2014a; Kocher et al. 2015). The finding that Cimicoidea is not monophyletic is unexpected, though this research is the first to investigate relationships within this superfamily (Schuh et al. 2009; Wierauch and Schuh 2010; Kocher et al. 2015). Obtaining sequence data from more species within this clade will shed light on this finding. Considering the present study and the recent data from Balvin et al. (2015), there is good reason to believe *Oeciacus* should be considered as synonymous with *Cimex*. As more sequence data of species within family Cimicidae becomes available, these relationships can be further parsed out (Balvin et al. 2015).

Literature Cited

Adelman, Z.N., Kilcullen, K.A., Koganemaru, R., Anderson, M.A.E., Anderson, T.D. and Miller, D.M. (2011) Deep sequencing of pyrethroid-resistant bed bugs reveals multiple mechanisms of resistance within a single population. PLoS ONE 6: e26228.

Axtell, R.C. (1999) Poultry integrated pest management: status and future. Integrated Pest Management Reviews 4: 53–73.

Bai, X., Mamidala, P., Rajarapu, S.P., Jones, S.C. and Mittapalli, O. (2011) Transcriptomics of the bed bug (*Cimex lectularius*). PLoS One 6: e16336.

Balvin, O., Munclinger, P., Kratochvil, L., and Jitka, V. (2012) Mitochondrial DNA and morphology show independent evolutionary histories of bedbug *Cimex lectularius* (Heterpotera: Cimicidae) on bats and humans. Parasitology Research 111: 457-469.

Balvin, O., Roth, S., and Vilimova, J. (2015). Molecular evidence places the swallow bug genus *Oeciacus* Stål within the bat and bed bug genus *Cimex* Linnaeus (Heteroptera: Cimicidae). Systematic Entomology 40: 652-665.

Beckenbach, A. T. and Joy, J. B. (2009) Evolution of the mitochondrial genomes of gall midges (Diptera: Cecidomyiidae): rearrangement and severe truncation of tRNA genes. Genome Biology and Evolution 1: 278-287.

Bitam, I., Dittmar, K., Parola, P., Whiting, M. F. and Raoult, D. (2010) Fleas and flea-borne diseases. International Journal of Infectious Diseases 14: 667-676.

Booth, W., Saenz, V.L., Santagelo, R.G., Wang, C., Schal, C., and Vargo, E.L. (2012) Molecular markers reveal infestation dynamics of the bed Bug (Hemiptera: Cimicidae) within apartment buildings. Journal of Medical Entomology 49: 535-546.

Booth, W., Balvin, O., Vargo, E. L., Vilímová, J., and Schal, C. (2015) Host association drives genetic divergence in the bed bug, *Cimex lectularius*. Molecular Ecology 24: 980-992.

Brown, C. R., and Brown, M. B. (2005) Between-group transmission dynamics of the swallow bug, *Oeciacus vicarius*. Journal of Vector Ecology 30: 137.

Cameron, S.L. (2014a) Insect mitochondrial genomics: implications for evolution and phylogeny. Annual Review of Entomology 59: 95-117.

Cameron, S. L. (2014b). How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. Systematic Entomology 39: 400-411.

Cheng, S., Fockler, C., Barnes, W. and Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. Genetics 91: 5695-5699.

Clary, D. O. and Wolstenholme, D. R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. Journal of Molecular Evolution 22: 252-271.

Crozier, R. H. and Crozier, Y. C. (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics 133: 97-117.

Cui, Y., Xie, Q., Hua, J., Dang, K., Zhou, J., Liu, X., Wang, G., Yu, X., and Bu, W. (2013) Phylogenomics of Hemiptera (Insecta: Paraneoptera) based on mitochondrial genomes. Systemics Entomology 38: 233-245.

Davies, P. A. and Gray, G. (2002) Long-range PCR, in Methods in Molecular Biology, PCR mutation detection protocols 187: 51-56.

Davies, T. G. E., Field, L. M., and Williamson, M. S. (2012) The re-emergence of the bed bug as a nuisance pest: implications of resistance to the pyrethroid insecticides. Medical and Veterinary Entomology 26: 241-254.

Delaunay, P., Blanc, V., Giudice, P.D., Levy-Bencheton, A., Chosidow, O., Marty, P. and Brouqui, P. (2011) Bed bugs and infectious Diseases. Clinical Infectious Diseases 52: 200-210.

Doggett, S.L., Geary, M.J. and Russell, R.C. (2003) Has the tropical bed bug, *Cimex hemipterus* (Hemiptera: Cimicidae), invaded Australia? Environmental Health 3: 80–82

Dogget, S.L. (2007) Bed bugs-latest trends and developments. Synopsis of the Australian Environmental Pest Manager's Association National Conference, 4-6 July, Pacific Bay Resort, Coffs Harbour, pp. 22-37

Dogget, S. L., Dwyer, D. E., Penas, F. P. and Russel, R. C. (2012) Bed bugs: clinical relevance and control options. Clinical Microbiology Reviews 25: 164-192.

Dotson, E. M. and Beard, C. B. (2001) Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. Insect Molecular Biology 10: 205-215.

Fenn, J. D., Song, H., Camreon, S. L., Whiting, M. F. (2008) A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. Molecular Phylogenetics and Evolution 49: 59-68.

Fountain, T., Duvaux, L., Horsburgh, G., Reinhardt, K., and Butlin, R. K. (2014) Human-facilitated metapopulation dynamics in an emerging pest species, *Cimex lectularius*. Molecular ecology 23: 1071-1084.

Francischetti, I. M., Calvo, E., Andersen, J. F., Pham, V. M., Favreau, A. J., Barbian, K. D., Romero, A., Valenzuela, J. G., and Ribeiro, J. M. (2010) Insight into the sialome of the bed bug, *Cimex lectularius*. Journal of Proteome Research 9: 3820-3831.

Gries, R., Britton, R., Holmes, M., Zhai, H., Draper, J., and Gries, G. (2015) Bed bug aggregation pheromone finally identified. Angewandte Chemie 127: 1151-1154.

Henry, T. J. (2009) Biodiversity of Heteroptera. Insect biodiversity: science and society: 223-263.

How, Y. F. and Lee, C.Y. (2009) Survey of bed bugs in infested premises in Malaysia and Singapore. Journal of Vector Ecology 35: 89–94.

Hu, M., Chilton, N. B. and Gasser, R.B. (2002) Long PCR-based amplification of the entire mitochondrial genome from single parasitic nematodes. Molecular and Cellular Probes 16: 261-267.

Jones, N. L. (2002) PCR: Principles, procedures and parameters, in Methods in Molecular Biology, PCR mutation detection protocols. 187: 37-46.

Jung, S. and Lee, S. (2012) Correlated evolution and Bayesian divergence time estimates of the Cimicoidea (Heteroptera: Cimicomorpha) reveal the evolutionary history. Systematic Entomology 37: 22-31.

Keller, I., Benasson, D., and Nichols, R. A. (2007) Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. PLoS Genetics 3(2): e22. doi:10.1371/journal.pgen.0030022

Kim, W. and Yeung, E. (1997) Separation of DNA sequencing fragments up to 1000 bases by using poly(ethyleneoxide)-filled capillary electrophoresis. Journal of Chromatography 781: 315-325.

Kocher, A., Kamilari, M., Guilbert, E., Lhuillier, E., Coissac, E., Peneau, J., Chave, J. and Murienne, J. (2014) Shotgun assembly of the assassin bug *Brontsoma colossus* mitochondrial genome (Heteroptera, Reduviidae). Gene 552: 184-194.

Kocher, A., Guilbert, E., Lhuillier and E., Murienne, J. (2015) Sequencing of the mitochondrial genome of the avocado lace bug *Pseudacysta perseae* (Heteroptera, Tingidae) using a genome skimming approach. Comptes Rendus Biologies 338: 149–160.

Laslett, D. and Canbäck, B. (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24:172-175.

Li, H., Gao, J., Liu, H., Liang, A., Zhou, X. and Cai, W. (2011) The architecture and complete sequence of mitochondrial genome of an assassin bug *Agriosphodrus dohrni* (Hemiptera: Reduviidae). International Journal of Biological Sciences 7: 792-804.

Li, H., Liu, H., Cao, L., Shi, A., Yang, H. and Cai, W. (2012a) The complete mitochondrial genome of the damsel bug *Alleorhynchus bakeri* (Hemiptera: Nabidae) International Journal of Biological Sciences 8: 93-107.

Li, H., Liu, H., Shi, A., Stys, P., Zhou, X. and Cai, W. (2012b) The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). PLoS ONE 7(1): e29419.

Li, T., Gao, C., Cui, Y., Xie, Q. and Bu, W. (2013) The complete mitochondrial genome of the stalk-eyed bug *Chauliops fallax* Scott, and the monophyly of Malcidae (Hemiptera: Heteroptera). PLoS ONE 8: e55381.

Lowe, C.F. and Romney, M.G. (2011) Bed bugs as vectors for drug-resistant bacteria. Emerging Infectious Disease 17: 1132.

Ma, C., Yang, P., Jiang, F., Chapuis, M. P., Shali, Y., Sword, G. A., and Kang, L. E. (2012) Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. Molecular ecology 21: 4344-4358.

Moriyama, M., Koga, R., Hosokawa, T., Nikoh, N., Futahashi, R. and Fukatsu, T. (2012) Comparative transcriptomics of the bacteriome and spermalege of the bedbug *Cimex lectularius* (Hemiptera: Cimicidae). Applied Entomology and Zoology 47: 233-243.

Ng, P. and Henikoff, S. (2006) Predicting the effects of amino acid substitutions on protein function. Annual Review of Genomics and Human Genetics 7: 61-80.

Panagiotakopulu, E. and Buckland, P. (1999) *Cimex lectularius* L., the common bed bug from Pharonic Egypt. Antiquity 73: 908-911.

Pfiester, M., Koehler, P. G., and Pereira, R. M. (2009) Effect of population structure and size on aggregation behavior of *Cimex lectularius* (Hemiptera: Cimicidae). Journal of medical entomology 46: 1015-1020.

Polanco, A.M., Brewster, C.C. and Miller, D.M. (2011a) Population growth potential of the bed bug, *Cimex lectularius* L.: a life table analysis. Insects 2: 173–185.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardaso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D. and Vogler, A. P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55: 595-609.

Potter, M. F., Rosenberg, B., and Henriksen, M. (2010) Bugs without borders: defining the global bed bug resurgence. Pest World 18: 8-20.

Qu, J., Richards, S., Bandaranaike, D., Bellair, M., Blankenburg, K., Chao, H., Dinh, H., Doddapaneni, H., Downs, B., Dugan-Rocha, S., Elkadiri, S., Gnanaolivu,R., Hernandez,B., Javaid, M., Jayaseelan, J.C., Lee, S., Li, M., Ming, W., Munidasa, M., Muniz, J., Nguyen, L., Ongeri, F., Osuji, N., Pu, L.L., Puazo, M., Qu, C., Quiroz, J., Raj, R., Weissenberger, G., Xin,

Y., Zou, X., Han, Y., Worley, K., Muzny, D. and Gibbs, R. (2015) *Cimex lectularius* genome sequencing. Unpublished. GenBank: KK244448.1

Raupach, M.J., Hendrich, L., Kuchler, SM., Deister, F., Moriniere, J. (2014) Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. PLoS ONE 9: e106940

Reeves, W. K., Loftis, A. D., Gore, J. A., and Dasch, G. A. (2005) Molecular evidence for novel Bartonella species in Trichobius major (Diptera: Streblidae) and *Cimex adjunctus* (Hemiptera: Cimicidae) from two southeastern bat caves, USA. Journal of Vector Ecology 30: 339.

Reinhardt, K., and Siva-Jothy, M.T. (2007) Biology of the bed bugs (Cimicidae). Annual Review of Entomology 52: 351-374.

Richards, L., Boase, C.J., Gezan, S. and Cameron, M.M. (2009) Are bed bug infestations on the increase within Greater London? Journal of Environmental Health Research 9: 17–24.

Romero, A. (2011a) Biology and management of bed bugs: review of recent research. Proceedings of the Seventh International Conference on Urban Pests: 307-310.

Romero, A. (2011b) Moving from the old to the new: insecticide research on bed bugs since the resurgence. Insects 2: 210-217.

Roux, K. H. (2009). Optimization and troubleshooting in PCR. Cold Spring Harbor Protocols 4: pdb-ip66.

Saenz, V.L., Booth, W., Schal, C., and Vargo, E.L. (2012) Genetic analysis of bed bug populations reveals small propagule size within individual infestations but high genetic diversity across infestations from the Eastern United States. Journal of Medical Entomology 49: 865-875.

Salazar, R., Castillo-Neyra, R., Tustin, A. W., Borrini-Mayorí, K., Náquira, C., and Levy, M. Z. (2014) Bed Bugs (*Cimex lectularius*) as vectors of *Trypanosoma cruzi*. The American journal of tropical medicine and hygiene. 14-0483.

Seong, K.M., Lee, D.-Y., Yoon, K.S. (2010) Establishment of quantitative sequencing and filter contact bioassay for monitoring pyrethroid resistance in the common bed bug, *Cimex lectularius*. Journal of Medical Entomology 47: 592–599.

Schaafsma, E. J., Hapke, S. D., and Banfield, M. G. (2012) Bed bug (*Cimex lectularius* L.) population composition as determined by baited traps. Insects 3: 442-452.

Schuh, R. T., Weirauch, C., and Wheeler, W. C. (2009) Phylogenetic relationships within the Cimicomorpha (Hemiptera: Heteroptera): a total-evidence analysis. Systematic Entomology 34: 15-48.

Simon, S. and Hadrys, H. (2013) A comparative analysis of complete mitochondrial genomes among Hexapoda. Molecular Phylogenetics and Evolution 69: 393–403.

Stork, N.E. (1988) Insect diversity: facts, fiction and speculation. Biological Journal of the Linnean Society 35: 321-337.

Stutt, A. D. and Siva-Jothy, M. T. (2001) Traumatic insemination and sexual conflict in the bed bug *Cimex lectularius*. Proceedings of the National Academy of Sciences 98: 5683-5687.
Swofford, D. L. (2001) Paup*: Phylogenetic analysis using parsimony (and other methods) 4.0. B5.

Swofford, D. L. (2002) PAUP* version 4.0 b10. Sinauer, Sunderland, MA.

Szalanski, A. L., Austin, J. W., McKern, J.A., Steelman, C.D., and Gold, R. E. (2008) Mitochondrial and ribosomal internal transcribed spacer 1 diversity of *Cimex lectularius* (Hemiptera: Cimicidae). Jounral of Medical Entomology 45: 229-236.

Szalanski, A.L., Tripodi, A.D. and Austin, J.W. (2011) Multiplex polymerase chain reaction diagnostics of bed bug (Hemiptera: Cimicidae). Journal of Medical Entomology 48: 937-940.

Usinger, R.L. *Monograph of Cimicidae (Hemiptera-Heteroptera)*; Thomas Say Foundation: College Park, MD, USA, 1966.

Vargo, E. L., Booth, W., Saenz, V., Santangelo, R.G., and Schal, C. (2011) Genetic analysis of bed bug infestations and populations. Proceedings of the Seventh International Conference on Urban Pests.

Wang, C., Gibb, T., Bennet, G. W., and McKnight, S. (2009) Bed bug (Heteroptera: Cimicidae) attraction to pitfall traps baited with carbon dioxide, heat, and chemical Lure. Journal of Economic Entomology 102: 1580-1585.

Wang, C., Saltzmann, K., Chin, E., Bennett, G. W., and Gibb, T. (2010) Characteristics of *Cimex lectularius* (Hemiptera: Cimicidae), infestation and dispersal in a high-rise apartment building. Journal of Economic Entomology 103: 172-177.

Weirauch, C. and Schuh, R. T. (2011) Systematics and evolution of heteroptera: 25 Years of Progress. Annual Review of Entomolgy, 56: 487-510.

Williams, K. and Willis, M. S. (2012) Bedbugs in the 21st century: the reemergence of an old foe. Lab Medicine 43: 141-148.

Yang, W., Yu, W. and Du, Y. (2013) The complete mitochondrial genome of the sycamore lace bug *Corythucha ciliate* (Hemiptera: Tingidae). Gene 532: 27-40.

Yang, Z. and Yoder, D. A. (1999) Estimation of the transition/transversion rate bias and species sampling. Journal of Molecular Evolution 48: 274-283

Yoon, K.S., Kwon, D.H., Strycharz, J.P., Hollingsworth, C.S., Lee, S.H. and Clark, J.M. (2008) Biochemical and molecular analysis of deltamethrin resistance in the common bed bug (Hemiptera: Cimicidae). Journal of Medical Entomology 45: 1092-1101.

Zhu, F., Wigginton, J., Romero, A., Moore, A., Ferguson, K., Palli, R., Potter, M.F., Haynes, K.F. and Palli, S.R. (2010) Widespread distribution of knockdown resistance mutations in the bed bug, *Cimex lectularius* (Hemiptera: Cimicidae), populations in the United States. Archives of Insect Biochemistry and Physiology 73: 245-257.

Zorrilla-Vaca, A., Silva-Medina, M. M., and Escandón-Vargas, K. (2015) Bedbugs, *Cimex* spp.: their current world resurgence and healthcare impact. Asian Pacific Journal of Tropical Disease 5: 342-352.

APPENDIX A

Nucleic acid sequence data of *Cimex lectularius* (5' TO 3')

TTAGGTTCATACCCTAAATATAGAAGCACTAAACTTCTTCTTTTTAATTTTTAAATATTCTAATAAA
ATACTATTTTTTACCATAATAATTTTAGGTTCTTTACTAGTATTAAGATCAAATAATTGATTAAGAA
TTTGAATAGGGTTAGAAATTAATCTAATATCTTTTATTCCGTTAATACTAAATTCTAAAAACAGAT
CACTCTCAGAAGGATGTATAATATATTTTTTAATCCAAACTATTAGCTCAATTTTATTATTATTTTC
TATTTTAATTAAGCCTGCACAAATTATTTTATTAATAAATATAACCCAAAATATTATTATTGTAAGA
ATAATAATAAAATTAGGCTTACCTCCATTTCACTATTGATTCGTATTTGTAATAAAAAAAGTAGAA
TGATATTCCTGCTTAACTCTTATAACCTGACAAAAAATTGCCCCCTTAACAATTTTATCGTACATAG
ACTTCAATATCATTATTGCATTTATTGCCTCCATTGTAGCAAATATTACTGCCATTAATCAAACAGA
TTTACGAAAAATTATAGCTCTGTCATCTATTAATCATTTAGGATGAATATTAACTTGCATAAAAGA
AAGTAACACATGAATAATTTACTTATTATTTTATTCTTTAATGTCAGCATCCTTATTAATCTTTCTTA
AAAAAAAGAAAATCCTTTTTATTAATCAAATAAATAAACTAAACAACAATAGTGATAAACTAACT
TTATCTATATTAATAATAAGAATAGCAGGATTACCTCCCTTTATCGGGTTTATGCCTAAATGAATA-
GCAATTCAATATATAATAATAAATATGAATATTTTATTAGCATATTTTATAGTAATAATATCCTTAA
TTGCAGTATTTTATTATTTACGCATAATTAACGCAACTTTAATGCAAAATTCCTGTAAAAATAAGT
GAACTTTAAACCAAAGAAAATCATATAGATCGCTTATATTAGTATGTAATTTATTTTTGCCCATCTC
AATTTTTCTAATTAATTAAAGTCTTAAGTTAAAAAAACTATTAACCTTCAAAGTTAAAAACATAAA
TATATTATAGACTTTAGCCTTACGCAACTTCAAACTTGCAATTTGACATCATAATTTTTTCTCTAAA
ATCCTAATAAGAGGCTCATGCCATAAATAAATTTACAATTTATTGCCTTAAATTCAGCCACCTTAT
TAATGGTTAAGTGAATAATGTCGACCAACCACAAAGACATCGGAACTTTATATTTTTTATTCGGGA
TATGAGCAGGTATACTAGGAACATCAATAAGATGAATTATTCGAATTGAACTTAGACAACCTGGC
TCATTTATTGGAGATGATCAAATTTATAATGTAATTGTTACTGCTCATGCTTTTGTAATAATTTTTTT
CATAGTTATACCAATTATAATTGGGGGATTTGGGAATTGACTAGTGCCATTGATAATTGGCGCACC
TGATATAGCATTCCCTCGACTCAATAATATAAGATTTTGACTTGTACCTCCTTCCTTATTACTTTTA
TTAGTCAGAAGAACATCGAGAACAGGTGTAGGGACAGGATGGACAGTTTATCCCCCCTTATCGGG
TAACATTGCTCACATAGGATATTCTGTTGACTTTGCGATCTTCAGATTACATCTCGCAGGAATAAG
ATCTATTTTAGGGGCAATCAACTTTATCTCCGCTATTTTAAATATACGTCCTGCGGGGATAACCTTA
GAACGAACTCCCTTATTTGTATGATCCGTAGGAATCACAGCAATACTACTACTACTTTCCCTACCA
GTACTTGCTGGAGCAATCACCATATTATTAACTGATCGTAACTTCAATACTTCATTTTTTGATCCAG
TAGGGGGGGGAGATCCTGTACTTTACCAACACTTATTTTGATTCTTCGGCCACCCTGAAGTATATA
TTCTAATTCTACCAGGATTTGGTTTAATTTCACATATTATCAGCAAAGAAAGAGGAAAAAATGAAA
CCTTTGGTCCTTTAGGTATAATCTATGCTATATCGGCAATCGGGATCCTAGGATTTATTGTTTGGGC
ACATCACATATTTACAGTAGGAATAGATGTAGACACTCGAGCATATTTCACCTCTGCAACTATAAT
TATTGCTGTACCAACAGGTATCAAAATTTTCAGATGAATCGCTACACTATATGGAAGAAAAATGTC
ATTCTCACCATCTATTATATGAGCAATAGGATTTGTATTTTTATTTACCATTGGCGGTATAACAGGG
GTAATTTTAGCTAATTCATCTATTGACGTAGCATTACATGATACTTATTACGTAGTAGCTCATTTCC
ATTATGTACTTTCAATAGGAGCCGTATTTGCTATTATGGGAAGATTTATTCAATGATTCCCTCTTTT
TACGGGATTAACAATAAACCCAAAATGATTAAAAATTCAATTCTTAACCATATTTATCGGTGTAAA
CATAACCTTCTTCCCGCAACACTTCTTAGGATTAAACGGGATACCGCGTCGGTATTCAGACTATCC
TGATATATATATATCTTGAAATGTACTGTCATCTCTAGGTAGAACCATCTCAATTATAAGTATTTTC
ATATTCATAATAATCATCTGAGAGGCTCTTGTATCAAAACGTAAAAGCCTGTTTAATAATAATATG
CCTTCCAATATTGAATGATTACAACTTTCCCCTCCCAGAGAACACTCTTATAACGAGTTACCAATT
ATTTGCTCATTCTAATATGGCAGATTAGTGCAATGAACTTAAGATTCATTTATAAAGGCTTACCTTT
TATTAGAAATGCCAATCTGAAATTCATATTCATTTCAAGATGCAAACTCTCCCATTATAGAACAAT
TAACTTTTTTCCATGATAGCACAATAATAATTCTTACAGCAATTATTACTATTATTTCATATCTACT
GATTTACATCACATTTAATAAGCTAACTAATCGATATCTTTTAAATAGACAAAATATTGAACTATT
ATGAACAATTACGCCAGCAGTAACATTATTATTCATTGCATTACCTTCATTAAAAATTCTATATCTT
ATAGACGACATCACTAATCCAATAATAACAATTAAAGCTATTGGACACCAATGATATTGAAGCTA
CGAATATTCTGATTTCAAAAATATCGAATTCGAATCTTATATAAAATTGCCTTCAGAAATAAATGA
AAACGAATTTCGACTGCTAGAAACTAATAATCGTACTATCTTACCAATTCATACTTCAATCCGATT

GTTGACTACTTCCACAGACGTGATTCATTCATGAACAGTTCCATCTTTAGGAGTTAAGATTGATGC
AATCCCAGGACGATTAAATCAAAGATTAATTATTATTAATCGACCAGGATTAATATATGGACAAT
GCTCTGAAATCTGTGGAGCAAATCATAGATTTATGCCTATTGTCCTAGAAAGAATCAGTATTCAAA
GATTCATTAAATGAATAAATTTATACTCATTAAGTGGCTGAAATGTTAAGCAATGGTCTCTTAAAC
CACTAAATAGTAAATTCACAAATACTCTTAATGAAAGAATTAGTTTCAAAAAAAACAATAACTTG
TCAAGTTATAAATACTTAATAGTATTCTTTAATTCCACAAATATCACCAATATGATGAACTACACT
AATAACAATCTTTATTTTATGTACATTTATATTTTCAATTAACTTATTCTTTTGTTTTAAAACAAAAT
TAACTTTTAATAAAAATAAATCCCAAAATAATATAAATTGAAAATGATAGCTAATTTATTCTCTTC
ATTTGACCCATCAACTAGAATTATTTCGTCATTAAATTGATATAGAACTATTATTGGTCTAATAATA
ATCCCGCCAATATACTGAACTACGCCTTCACGAATTACTATATTATTTATTTTATTATATTATAAAC
TATACACAGAATTTAAAATTACAATAAGAAAACCAAATGCCAAAATAAACTTAATAATAATTTCA
ATATTTATATTCATTTTAACAAATAATATAATAGGGTTATTACCTTATGTATTTACTAGCTCCAGTC
ATATAGTATATTCACTATCATTGTCATTACCATTATGATCAGCCCTAATAATATACACATGAATTA
ATTTTACAAACAAAACATTCACACATATACTCCCTAATGGCACTCCTGCCTTACTAATGCCGTTTTT
AGTATGCATTGAAACAGTAAGAAATATTATTCGACCTATATCCTTAGCAATCCGGCTAACAGCAA
ATATAATTGCAGGCCACTTATTCATAGTATTATTAGGAAATTTAATTGTTCATAGAGAACATTATA
TGTTAATTGTAATTACACAAACATTGCTAATAATATTTGAATTATGTGTATCATTTATTCAAGCTTA
CGTATTCACTACATTAAGCACATTATACATTAAAGAAGTAGCATATGAAACAAGACAAAAACCAC
CCATATCATATAGTAGATTATAGCCCTTGGCCTTTAACTGGATCAATTGGAGCATTAACATTAACC
TTCGGAATAGTTATTTGATTTCATAAAAAAGAAAGAATATTAATAATATTAGGAATAGTTATTATT
ATTTTAACTATATTTCAATGATGACGAGATATTATTCGAGAAAGAACATATCAAGGACATCATACA
TTAAAAGTCGCCAAAGGAATTAAAATAGGAATAATCTTATTCATTATCTCAGAAATTTTATTCTTT
ATATCATTTTTCTGAGCTTTTCTGCATAGAAGAATTGCTCCTTCCATTGAAATTGGAAATAACTGAC
CCCCTTTGGGGATTAAAACGTTTAACCCTATAGAAATCCCATTATTAAATACAATAATCCTATTAA
GTTCAGGGGTAAGAGTCACATGGGCCCACCACAGAATTATAGAAAGTAAACATAAACAAACCAC
AAATAGTCTAACTATCACTGTAATATTAGGAATTTATTTCTCGTTATTACAAGGATATGAATATAA
TCAAGCTGAATTCACTATTGCAGACTCTGTATATGGATCTTGTTTCTTTATGATTACAGGATTTCAT
GGATTTCATGTTATAGTTGGAACTATTTTCTTAATTGTATGCTTACTACGAAATATAAAAAATCACT
TCTCCAGCCTCATGATACTGACACTTCGTAGATGTAGTATGACTATTCGTTTATATTATCATTTATT
GATGAGGAAGATAGCTTATTCTTTTAGTATATTAGTATATCTGACTTCCAATCAGAAGGACTTTAA
TTAGAAAGAATAATAAAAATAATTTTAACCACATCCTTCATGACTTTAACCATTTCAGTACTAATA
ATAATAATTTGCATTTTTACAAGAAAAAAAAGTATCATAGATCGAGAAAAAATATCACCATTCGA
ATGTGGGTTTGACCCTAAAAGATCACCTCGTATACCATTCTCAATTCAGTTCTTTCTAGTAGCAAC
ATTATTTCTAGTATTTGATGTAGAAATTACCATTATCATTCCTATAATTATTACATTTAAAATATCT
TCCACAAGAACATGACTCTTTATAACAACAACATTTATTATCATTTTACTGCTGGGATTATATTACG
AATGATATAATAACATACTTGAATGATCTTTCTAGGGAAAGTAGTTAACTATAACATTTAATTTGC
AATTAAAAAGTACTATATCGTAGTCTTTCTCATTTAAGTAATGAAGTAAAATTTACATTTAGTTTCG
ACCTAAAGACAAGAGAACTTATTTCTCCTTACTTTTAATAGAAGCCAAAAAGAGGCAATCTATTGT
TAATAGAAAAAATGACAAAATTGTCCTATTAAAAGAGAATGATTGTAACTAAAAGAAGCTGCTAA
CTATCTTTTAAAGCGGTTAAAATCCGTTTTTCTCTTATTTAAGTAGTTTATAAAAAACATTTCATTT
TCAGTGAAAAGAAAATAATAATTCTTAAATATCAAAAGAGATATTAGTCTCATACTAACAACTTC
AATGTTATGCTTTAAATTTAAGCTATTTTAATTAAAGAAAGAAAAAGAAAGATCTTAAAATAAAA
GAAATTAAAATTATGCTAATTCCTTTACTATAAAACCCTGAAACATTAACTGAAACCCTGATTAAA
AAATGAAAAATCGATCGCGAAACAATATATTCCCCCCAACTACAATCAATATTTGTAATCCAAAC
CTTAGGAATATAAAATCCAAAACTGCTCAAAAGAAAATTTCTGAAATATGGCATATAAAGTATAA
ACCCAAAAAATAAATGAAACAAATAGAAAGAAGAACCCCAATTTATTGAACATATTTTATAAGAA
ATCAGAATCCCTATTAAAATTAACAAAAAAGGTGCTAACTTAGTATATAAATGCAAACAATAAAA
TTCAGGAACCGTAAAAATTACTCAACTTAAAAAGCTACCCCCTAAAACGGATATTATAGCCATAA
ATAGTATTGAAAAAATTATTGAATTATCCTCAAAGTAACTTTGACAGCAATATAAACTCTCTCTGC
CAATAGTATAATAAATTAAACGAACAGAATATGAAACTGTTAACCCTACAGATAAATAATACATT
AAGACAATAAAAAAATTGTCGTAATTAAATCTAATAACCTCCAAAATCATATCCTTAGAATAAAA
CCCTGATATAAAAGGGATACCACATATAGATATACTAGAAATTAAAAACGAAGTATAAGTTCAAG

GAATACATTTATAAATGCCCCCTATAACACGTAAATCTTGATTATTACCTGCAATATGAATAATTA
AGCCTGCGCATAAAAATAGTAAAGCCTTAAAAAAAGCATGTATTAATAAATGAAAAAATCTTAAA
ACAGGATAACCTATAAATAAAACTCCTATTATAAAACCTAATTGACTCAAAGTTGATAAAGCAAT
TACCCTTTTTAAATCAAATTCAAATATAGCCCCCAACCCAGCAACAAACATCGTTATTGTACCAAT
ATAAACAAAAAATTCGCAATTTAAATTATTAAATAAATATCTAAAACGAATTAATAAATAAACCC
CAGCAGTTACTAAAGTAGATGAGTGAACTAAAGCAGAAACAGGAGTTGGGGCCGCTATAGCAGC
AGGTAATCAGGAAGAAAACGGAATTTGAGCTTTTTGTAAACTAGCTAAAATTACTATCAACAAAA
CTATCTTTATATTAAAATCTCAAACATTAAAATAAAACAAATAATGTCAACCTCCAAAATTAAATA
TTCAAGCTACAGCCATTAATAAAGCAGCATCACCGATTCGATTAATTAAAATTGTTAACATGCCCG
CATTATAAGACTTAAAATTTTGAAAATAGATAACTAAACCATAAGAAACTAAACCAAGACCATCT
CAACCTAAAAGAATTCTAATTAATCTGGGACTTACAATTATAAGTAATATTCTCAAAACAAACAAT
AAAACCAAATATAAAAATCGAATATTATTGCTATCCTCTGATATATAAGAAAATCTATATAAAATA
ACCAAAGAAGAAATAATTAAAACAAAAGATATAAATATTAAAGATATTCAATCCAATAATAAAAC
TGCAGTAAAGCTACATGAATTAAGAGAAAAAATTTCCCAATCTATAAAAATAATATAACCGTAAT
TAACAAAACAAAGGCCTACAACAAATAAACTCAAACCAATCACAAATAAATAAAAAGAACTTAA
TAAATACAATATCACTTACAGTATAAAAATACGCCCTTAGTTCCACAAACTAATATTCTATGCTAA
ACTATAATAAGTTATAAAAAAACAAAGAGAAAATCCACCTTAAAAACAAAGAAATTCAAAGGAA
ACCAATGTATAAAAATTAATAAAAAATCCCGAATTCTAGCACTAGAACCAAAAAACCCACCTTTA
TAAAAAACCCCGTGCTGAGTTATGCAATATAAATAGATTCTGTAACAACAACTTATAAAAGAATT
AATCATTAAAAACACCATAGCTAAATTTCTTCAACTTATAATACTATTCAACAATAATACTTCACC
TAACAAATTTAAAGAAGGAGGACTAGATATATTATTAGCTCTAAATATAAATCAAAACAAACTTA
TAGAAGGCATAATTGAAATAAAACCCTTATTAATAAACAATCTTCGTCTTAAAGAACGTTCATATA
TTACATTAGCTAAAACAAATAATCCAGAAGAACACAACCCATGACCTAATATTAAAACAAAAGAA
CCTATTAATCCTGACAAATTAAAAGTAAACAAACCACAAATAACTAAACCTATATGAGCAACTGA
AGAATAAGCAATTAAGCATTTAATATCAATTTGAAATAAACAAACAAAACTAGCAATTACAGAAC
CTAACAATCCTAAACTAATTATAATATAAGAAACATTGACATAATATGTATACAAAAAAAATATA
ATACGAATTAAACCATAACCGCCTAATTTTAATAAAACACCAGCTAAAATTATAGAACCAAAAAC
AGGAGCCTCAACATGAGCCCTGGGCAATCAAAAGTGAAAAATAAACATAGGCATTTTAACCAAAA
AAGCTAAAATCATAGACAAATGTAAATAAAAATTATTAGCTCCTTCTAAAAATCATCAAACAAAT
CTTCCTGACATAGAACCCAAATAAACAATGCCTAATAAAAGGGGCAAGGAAGCAAAAAGCATAT
AAAACAATAAATAAAACCCCGCATTCAACCGCTCTGGTTGGTACCCTCAGCCAAAAATTAAAAAC
AAAGTAGGAATCAATGTAGCCTCAAAATAAACATAAAATATAAAAACATTTATAGTTAAAAAACA
CAAAACCAAAAATAAAAGTAAAGATATACAAATAATCAAGAAATCGGCTATAGTATCTATATGTT
TCATGCAAAACCCTGAAAGAATTATTAAAACTACAATTCACAAAGACAAGTAAACCAAAAACCAC
GAAAAAATATCCCCACCAAAAGAGAAAGACAATGTACTAAAAAATCCTAAAAACCAATTTCTAAA
AAATCTTACAAAAAATAAAATTAACAAACAACCTATTAATTCGCAATAAAGCCCTAATATAGCAA
TAGGAATCAAAAATAACAAAGATAAAATTATTATCATATTAACATGGACATGGTTATCAAATTAT
CATTACCGTGACAACGAATTAAACTAACCAAAAGAGAAAGACCCAAAACCCCCTCACAAACAGA
AAAAGACAAAAAGACCAAAATGAAGTAATAACCAGATCCTAACCCTGTAAAAAAAACAAAAAAC
ATTAAAAACAAAATAAGAACCAAATATTCCAATCTTAACAATATCAAAAGTAAATGCCTATAAGC
TCTACAATAAGCAAAAAGACCAGATATAAATAAATACAATAAAAATCCAAGAATAAGTTTTAATA
GTTTAAAAAAAATAATGATCTTGTAAATCATAGATAGGGAAACCTTTAAAACTTCAACAAAAGAC
GAATTTATGTCTAACATTAATCTCCAAAACTAATATTTTCAATTAAACTACTTGTTGGCAGCTAGTA
AACAAAATAAATAAAGAGAAATATAATTTCCCCCCATAGGAGGGAAAAAATGGGGGGGATGTGA
AATATATAAGTAAACAACAGCTTGTTTAGCTGCTAACTTTAAGCAAGTGGTTTTAAAACCCACAAA
TAGGACATAAAACCTTTAAAGTTTAAGTATAAAGGGGTTAAAGCCTTACTAATTATTATTATAATT
ATTATAATGTTAATTTCAATCTACTTTATTAAATCAAAACACCCTTTAAGAATGGGAATTACTTTAA
TCTTTCAATCTATCTGTATTAGAATACTAACTGGAATAATCAGCTCATTTCTATTTTCTTATATTAT
AATAATAATTATATTAAGAGGGATATTAGTTTTATTTATATATATATCAAATGTGGCCTCTAATGA
AAAATTCTATATAACAAGCAAAAACTCATTATTAATTTTATTAATTCCACTTCTCATAATAAAACA
AGACTCAATTATATACAATTTACATAATTATAAAACTAATTCCACAATAAAATATGATATTTCCAC
CTCCTTAAGAAAGCTGTTTTCATCAGAAACAGTACCTTTAATTATAATAATAGTTTTATATTTATTA

TTCACTATAATTGTTATTTCATCAATTGTAAATATTCATGAAGGGCCAATACGCTCGAAAAGATAA
TGAATATATCAACACGTAAAACTAACCCTCTAATTAAAACATTAAACAATTTACTTATTGATCTCC
CATGTCCTTCAAGAATCTCCAATTGATGAAATTTTGGATCCTTACTAAGAATATGTTTGTTAATCCA
ATTATTAACAGGAATCTTTTTAGCTATACATTATACAGCTAACATTGAATTAGCCTTCAACAGTGT
AATTCACATTATACGTAATGTAAATAATGGTTGAATAATACGAAGTATACATGCTAACGGCGCCTC
ATTCTTTTTTATTTGTATATATATACATGTAGGGCGAGGAATTTACTATAATTCTTATCAACTAACT
AACACTTGAATAGTAGGAGTAATAATATTATTACTAACAATAGCAACAGCCTTCCTAGGTTATGTT
TTACCATGAGGACAAATGTCCCTGTGGGGGGCAACAGTAATCACAAACCTACTTTCTGTTATTCCA
TACGTAGGGAACAATATTATTCTATGGTTATGGGGAGGATTTTCTATTGAAAATGCCACATTAACA
CGATTTTTCGCTATACATTTCCTTATACCATTCATAATTACAGCTTTAGTGATAATTCATTTACTAT
ACTTGCACCAAACAGGCTCTAATAACCCGCTAGGGGTAGATAGTAATACTGATAAAATTCCATTTC
ACCCATATTTTTCTGTTAAAGATATAATAAGATGTATAATAATACTATTGGTTTTTATATTAATAAA
CACTATAGAACCTCAGCTTCTAGGAGATCCAGAAAACTTTATTCCAGCAAATCCATTAGTAACTCC
TGTTCACATCCAACCAGAATGATATTTTCTATTTGCATACGCAATTCTACGATCTATTCCTAATAAA
ATAGGAGGTGTAATCGCCATATTAGCAGCAATCAGAATACTAATAATTTTACCTTTGCTAAATAAG
CCCAATCTTCCTAATTCTAAATATTATCCCTTAAGTAAAATGTTATTTTGATTTTTTTGTTTTACTTG
ATTATTATTAACATGAATTGGCGCTAAACCGGCCGAAGAGCCATACATCTTTACAGGCCAAGCCCT
AA

Missing sequence of unknown size here

CCAGCCCCTCGAATTCAAGATAAAATAAAACATAATTCAGATTCTCCCTCAGCAAAATCGAAAGG
AGAGCGGTTAGTTTCTGCATATAATATATGTTAAATGCTCCAATTATTTTTATATTTTTTGCATTTTC
TTTTAAAGTCTATTAGCAGCTAGTAAACAAAATAAATAAAGAGAAATATAATTTCCCCCCATAAG
AGGGAAAAAATGGGGGGGATGTGAAATATATAAGTAAACAACAGCTTGTTTAGCTGCTAATTAGA
AAATTCTAATAATAATTAATAAATTAAAATAAATTAATAAATAATTTAATGAAACAGGTAAAAAA
CTCTTTCACGTTATATACATAAGTTTGTCATATCGATACCGGGGTAAAGTCCCTCGAACTCAAATA
AAACCTACACAAAATATTAAAACCTTAATATAAAACAACAAAGATCCTAAATCAGCCCCTCAAAA
CAAAATACAAGTAAAAGTTCTTATAAAAAGAATATTTATATATTCAGATAAAAAAATAAATGCAA
AACCCCCTCTTCTGTACTCAATATTAAACCCTGAAACTAATTCAGATTCTCCCTCAGCAAAATCGA
AAGGAGAGCGGTTGGTTTCTGCCAAACAAGAAATATATCAAGAGCCAAAAATAGGAAAAGAAAA
AAATACAAATCATACATCTTGCTGATAAGTTATAAAATCAATAAAATCAAATCCGCCTGTTAAAA
AGACCACTCCTAAGGCTAGTAAAGCCATTCTGACTTCATAGGAAATAGTCTGGGCCACAGAACGC
AGAGATCCTAATATAGCGTAATTAGAATTTGAAGATCAACCACTAATTATTACCCCATAAACCCCT
ACGCCTAAGCACACTATTATAAACAAAAACCCAAAACTAAAAAAGTATACATTTACTAAAAATGG
AAATAAACATCACAACATAAAAGATAAAATTATTATAAGTACTGGAGAACAATAATAAATAAAA
AAATTAGACAAACAGGGATATCTTTGTTCCTTAAAAAAAAGCTTTAAACCATCAGAAAAAGGCTG
TAATAATCCCATTAAGCCTACTTTATTAGGGCCCCTGCGAAGCTGAATATAGCCTAGAATTTTACG
TTCTATTAAAGTTACAAATGCCACGGCCACTAAAATACACAAAACCAAAAATAAAAAAGAAATCA
AGAACAATGCTATCTGTAATCTATATTACATGAATAAATTCTAAACTTATTGCACTAATCTGCCAA
AATAGCAATAATGATCAATTAAAAAAATTATTCCAACTACTCGGTCCTCTCGTACTAGAGTAATTA
ATACTAATTAAAGATAGAAACCGACCTGGCTCACGCCGGTCTGAACTCAGATCATGTAAAAAATT
AAAGGTCGAACAGACCTAATTATTAAGCGGCTACACCTAATATTAATTTTAATCCAACATCGAGGT
CGCAAACCTCGCCATCAATAAGAACTCTCCAACAAGATTACGCTGTTATCCCTAAGTAACTTAAAC
TTATAATCAAAACTAAATCTGGATCAAATAAACACTAATTAGTGAAAAACAAAAATAAAGTTAAA
TAAATTTACCTGTCGCCCCAACAAAACTCATTAAAAAAGTACACTATTAAAACCTAACAAAATTAT
ACTTCTATAATGAATAAAGTTTTAAAGGGTCTTATCGTCTACTAATGAAATTTCTGCTTTTTCACAG
AAAAATAAAATTAATTTAAGTAAGATAAAGAAAGTCAATCCCTCGTATGACCATTCATACTAGTTT
CCAATTAAAAAACAAATTATTATGCTACCTTTGCACAGTTAAAGTACTGCGGCCATTTAAAATTCA
TTGGGCAGGTCTGACCTTATATAAAAACAAAAAGACATGTTTTTGTTAAACAGGCGAAGATTTGAT
TTGCCGAATTACTATATCTTATTTTTAAAAATACTAATTAAATCATTTATACAATTTTATTTTCTATT
ACTAAAATAATCTAAATATAAAAATAAATATATACTAAATTATATGAACATTAAAATAAACTATA
ACGAAATTAATGATGGAAGTTTCTAATCCTACTAAAAATATATAAATTATAATATTATATATCATT

AACAAGCTAATCCCTGTTAATTTTTCAATCTTATATCACTACAAATGTAATTTTGATGAAATAAAA
TAGTCGAATTAAATTCGAATCTTTAGAAACCAGATATATTATGAACGAAAGACATATGACCTCTAT
ATTCTTATTCAAGATGTTATTGTAACAACAACCTCCATAATAATATATATCTCATAATTTCGGGATG
TATTAATATATTAAAATAATTTTGATAAACCCTGATACAAAAGGTACTCCAAATAAAGGATTCTTT
TCAGCAATGAAGAGATTTTCCTTCACAGTACTAATAAAATATAATTTATAAAAATGCTATTCAATA
AACTTTCTACTAAGAATATAATTATTTTTTGGAAAAAAAGGAAAAAAGAATTAATTATAAATCAAT
TATTTCAAATTAAACTGATTCAAACAGTTATACCATCGATGTAAATGATGGAACTACTCAGTAGTT
ATAATTTGTATAATCAAGATTTACTTTCCAGTAAATCTACTTTGTTACGACTTATCTTGAATCCAAG
AGCGACGGGCGATGTGTGCATAACTCAGAGCTTAAATCATAATTAATAAATATTAATTATTACTAC
CAAATCCTTATTCATTATAAATTTTCCATTAAAAAATCCTAAATTAATAATAAAGTAACCCACCAC
TAGCTTCTTTATTGCTGCACCTTGACCTGATATAGTTCTCATAAAGAAAGTAGAAAATAAACCTTA
TAATACATTCAATAACAGAGGTATACAAGCTTAAAAAAAGTAAAATTCCTCGTGGATCATCGATT
AAAGAGCAGGTCCTCTGGATAGACTAAGTTACCGCCAAATTCTTTTAATTTCAAGAGCATAACTAA
TACTACTAAGGTAAACTAAAATTACAAATTGAATAATAGGGTATCTAATCCTAGTTTCATGATCAA
TTCCTTAAATATTATTACCTAAAACTTTAATTATAATTAAGATTTCACCCAATAAATTATAATTATA
AGTTTTGCTGATACAAATATAATTTTTACCAAAATAATTAACTAGGACTTAATGAATAACCGCGAC
TGCTGGCACAAAATTAGTCAGACCATTACAAAAATTAACTAAGTCTTATTGACATAAATACTTAAT
TCTAAAAACTGCGACTAATTATAAAAGCTTCCCTTTAGAAAGCCTATTAACCAGCTTAAACATGCA
TGTATAATTATCTTGAAGCTAACTCCTAAGCTAGAATAAAACTTATATAGATTATAATATAAAAAT
TTGTCGGAGCATTATTAGTCCCCTCTCGAAAGGGATTTAGGGACCCCAGAATGGGGTATATTTATA
TATTCTTTTACTCCAATTATGGATTTTTTTTCTTCTACATACATAATTGGAGTAAAAGGAGGATATA
TAGAATATATTAAATTACAGTTTTAATAAAATAAAACTATATCCTTATTCATTAGCTTATATTTAAA
GTTGGATATAAAGTTTAATTAATTTGTGTTTTAAGCATAAATAATAAAATATAATAGCTTATAATA
ATATCAATTATAAAGTTTAATTAACTTGTGTTAGACAAAAAAAATACTATTAAA

APPENDIX B

Amino acid sequences of *Cimex lectularius*

ND2

MEALNFFFLIFKYSNKMLFFTMMILGSLLVLSSNNWLSIWMGLEINLMSFIPLMLNSKNSSLSEGCMMY
FLIQTISSILLLFSILIKPAQIILLMNMTQNIIIVSMMMKLGLPPFHYWFVFVMKKVEWYSCLTLMTWQKI
APLTILSYMDFNIIIAFIASIVANITAINQTDLRKIMALSSINHLGWMLTCMKESNTWMIYLLFYSLMSAS
LLIFLKKKKILFINQMNKLNNNSDKLTLSMLMMSMAGLPPFIGFMPKWMAIQYMMMNMNILLAYFM
VMMSLIAVFYYLRMINATLMQNSCKNKWTLNQSKSYSSLMLVCNLFLPISIFLIN

COI

MVKWMMSTNHKDIGTLYFLFGMWAGMLGTSMSWIIRIELSQPGSFIGDDQIYNVIVTAHAFVMIFFMV
MPIMIGGFGNWLVPLMIGAPDMAFPRLNNMSFWLVPPSLLLLLVSSTSSTGVGTGWTVYPPLSGNIAH
MGYSVDFAIFSLHLAGMSSILGAINFISAILNMRPAGMTLERTPLFVWSVGITAMLLLLSLPVLAGAITM
LLTDRNFNTSFFDPVGGGDPVLYQHLFWFFGHPEVYILILPGFGLISHIISKESGKNETFGPLGMIYAMSA
IGILGFIVWAHHMFTVGMDVDTRAYFTSATMIIAVPTGIKIFSWIATLYGSKMSFSPSIMWAMGFVFLFT
IGGMTGVILANSSIDVALHDTYYVVAHFHYVLSMGAVFAIMGSFIQWFPLFTGLTMNPKWLKIQFLTM
FIGVNMTFFPQHFLGLNGMPRRYSDYPDMYMSWNVLSSLGSTISIMSIFMFMMIIWEALVSKRKSLFNN
NMPSNIEWLQLSPPSEHSYNELPIICSF

COII

MPIWNSYSFQDANSPIMEQLTFFHDSTMMILTAIITIISYLLIYITFNKLTNRYLLNSQNIELLWTITPAVTL
LFIALPSLKILYLMDDITNPMMTIKAIGHQWYWSYEYSDFKNIEFESYMKLPSEMNENEFRLLETNNRTI
LPIHTSIRLLTTSTDVIHSWTVPSLGVKIDAIPGRLNQSLIIINRPGLMYGQCSEICGANHSFMPIVLESISIQ
SFIKWMNLYS

ATP8

MSPMWWTTLMTIFILCTFMFSINLFFCFKTKLTFNKNKSQNNMNWKW

ATP6

MSKPNAKMNLMMISMFMFILTNNMMGLLPYVFTSSSHMVYSLSLSLPLWSALMMYTWINFTNKTFT
HMLPNGTPALLMPFLVCIETVSNIIRPMSLAIRLTANMIAGHLFMVLLGNLIVHSEHYMLIVITQTLLMM
FELCVSFIQAYVFTTLSTLYIKEVAYETSQKPPMSYSSL

COIII

MKQDKNHPYHMVDYSPWPLTGSIGALTLTFGMVIWFHKKESMLMMLGMVIIILTMFQWWRDIIREST
YQGHHTLKVAKGIKMGMILFIISEILFFMSFFWAFLHSSIAPSIEIGNNWPPLGIKTFNPMEIPLLNTMILL
SSGVSVTWAHHSIMESKHKQTTNSLTITVMLGIYFSLLQGYEYNQAEFTIADSVYGSCFFMITGFHGFH
VMVGTIFLIVCLLRNMKNHFSSLM MLTLRSCSMTIRLYYHLLMSKMAYSFS

ND3

MKMILTTSFMTLTISVLMMMICIFTSKKSIMDREKMSPFECGFDPKSSPRMPFSIQFFLVATLFLVFDVEI
TIIIPMIITFKMSSTSTWLFMTTTFIIILLLGLYYEWYNNMLEWSF

ND5

MLYLLSSFYLFVIGLSLFVVGLCFVNYGYIIFMDWEIFSLNSCSFTAVLLLDWMSLMFMSFVLIISSLVIL
YSFSYMSEDSNNIRFLYLVLLFVLSMLLMIVSPSLISILLGWDGLGLVSYGLVIYFQNFKSYNAGMLTILI
NRIGDAALLMAVAWMFNFGGWHYLFYFNVWDFNMKMVLLMVILASFTKSAQIPFSSWLPAAMAAP
TPVSALVHSSTLVTAGVYLLIRFSYLFNNLNCEFFVYIGTMTMFVAGLGAMFEFDLKSVIALSTLSQLG
FMMGVLFMGYPVLSFFHLLMHAFFKALLFLCAGLIIHIAGNNQDLRVMGGIYKCIPWTYTSFLISSMSM
CGIPFMSGFYSKDMILEVISFNYDNFFIVLMYYLSVGLTVSYSVRLIYYTIGSESLYCCQSYFEDNSMIFS
MLFMAMMSVLGGSFLSWVIFTVPEFYCLHLYTKLAPFLLILMGILISYKMCSMNWGSSFYLFHLFFGF
MLYMPYFSNFLLSSFGFYIPKVWITNIDCSWGEYIVSRSIFHFLISVSVNVSGFYSKGISMILISFILSSFFFF
L

ND4
MMMILSLLFLIPIAMLGLYCELMGCLLILFFVSFFSNWFLGFFSTLSFSFGGDIFSWFLVYLSLWIVVLMI
LSGFCMKHMDTMADFLIICMSLLLFLVLCFLTMNVFMFYVYFEATLIPTLFLIFGWGYQPERLNAGFYL
LFYMLFASLPLLLGIVYLGSMSGSFVWWFLEGANNFYLHLSMILAFLVKMPMFIFHFWLPSAHVEAPV
FGSMILAGVLLKLGGYGLIRIMFFLYTYYVNVSYIMISLGLLGSVIASFVCLFQIDIKCLIAYSSVAHMGL
VICGLFTFNLSGLMGSFVLMLGHGLCSSGLFVLANVMYERSLSRSLFINKGFISIMPSMSLFWFMFSAN
NMSSPPSLNLLGEVLLLNSIMSWSNLAMVFLMINSFMSCCYSIYLYCMTQHGVFYKGGFFGSSASIRDF
LLIFMHWFPLNFFVFKVDFLFVFL

ND4l
MLDMNSSFVEVLKVSLSMIYKIIIFFKLLKLILGFLLYLFMSGLFAYCSAYSHLLLMLLSLEYLVLILFLM
FFVFFTGLGSGYYFILVFLSFSVCEGVLGLSLLVSLIRCHGNDNLMTMSMLMW

ND6
MKPLKFKYKGVKALLIIIMIIMMLISIYFIKSKHPLSMGITLIFQSICISMLTGMISSFLFSYIMMMIMLSGM
LVLFMYMSNVASNEKFYMTSKNSLLILLIPLLMMKQDSIMYNLHNYKTNSTMKYDISTSLSKLFSSETV
PLIMMMVLYLLFTMIVISSIVNIHEGPMRSKS

CytB
MNMSTRKTNPLIKTLNNLLIDLPCPSSISNWWNFGSLLSMCLLIQLLTGIFLAMHYTANIELAFNSVIHIM
RNVNNGWMMRSMHANGASFFFICM
YMHVGRGIYYNSYQLTNTWMVGVMMLLLTMATAFLGYVLPWGQMSLWGATVITNLLSVIPTDKIPF
HPYFSVKDMMSCMMMLLVFMLMNTMEPQLLGDPENFIPANPLVTPVHIQPEWYFLFAYAILRSIPNK
MGGVIAMLAAISMLMILPLLNKPNLPNSKYYPLSKMLFWFFCFTWLLLTWIGAKPAEEPYIFTGQALST
MYFMYFILSPLSSKFWDKMTQ

ND1
MALFLISFLFLVLCILVAVAFVTLMERKILGYIQLRSGPNKVGLMGLLQPFSDGLKLFFKEQSYPCLSNF
FIYYCSPVLMMILSFMLWCLFPFLVNVYFFSFGFLFMMVCLGVGVYGVMISGWSSNSNYAMLGSLRS
VAQTISYEVSMALLALGVVFLTGGFDFIDFMTYQQDVWFVFFSFPIFGSWYISCLAETNRSPFDFAEGES
ELVSGFNIEYSSGGFAFIFLSEYMNILFMSTFTCILFWGADLGSLLFYIKVLMFCVGFIWVRGTLPRYRY
DKLMYMTWKSFLPVSLNYLLIYFNLLIIISIF

Amino acids shown in grey were not used in phylogenetic anaylsis

APPENDIX C

Nucleic acid sequence data of *Cimex adjunctus* (5' to 3')

TAAAATATTATTTTTTATTATAATAATTATAAGAACNTNNTTAGTTTTAAGCTCAAACAATTGATTA
AGAATTTGAATGGGGTTAGAAATTAATTTAATGTCATTTATCCCTTTAATATTTAATCCTAAAAAT
ATGATAATCTCTGAAAGATGTATAATATATTTTTTAATTCAGAGAATTAGATCTATACTATTAATTT
TTTCTATTTTAATTAAACCTATACANNNACTCGAATTAATGAATATTACTCAGTATACAATTATAAT
TAGAATAATAATAAAGTTAGGATTGCCTCCATTTCATTATTGATTCGTAAATGTTATAAAAAAAAT
GGAATGATTATCATGCTTAGTGTTAATAACATGGCAAAAAATCGCCCCTATGATAATTATATCTTA
TGTGGATTTTAAGATTATTATTGCAATTATTGCCGCAATTGTAGCTAATATTACTGCTATCAATCAA
ACAGATTTACGAAAAATTATATCATTGTCGTCTATTAATCACATAGGATGAATACTAGCCTGCATA
AAAGAAAGAAACGCGTGGATCATTTATTTAATATTATATATATTAATATCAACACCATTATTAATT
TTTTTTAATAAAAAGAAAATTTTATTTATCAACCAGATAAATAAATTAAACACGACTAGAGACAA
ATTAACCTTATCAATGTTATTAATAAGAATGGCAGGATTGCCTCCATTTACAGGATTTTTACCAAA
ATGAATAGCAATTCAATATATAATCATAAACGAAAACATTTTTATAGCATTTTTAATAACAACAAT
GTCTTTAATCGCCGTATTTTACTATTTACGAATAATTAACACAATAATATTACTAAATTCAGATAA
ATCTAAATGAACCTTGTCTCATAATAAATCATATGGGTCTTATATGATAATATGTAATTTATTATTA
CCAATTTCACTATTTTTAATTGACTAAAGTTTTAAGTTATGATAAACTATTAACCTTCAAAGTTAAA
AATATAAGTAAATTATAGACTTTAGCTATAGCAACCTCAAATTTGCAATTTGATATCATAACTTTG
ACTATAAAATCCTAATAAGGGGCAAATGCCATAAATAAATTTACAATTTATTACCTAAAAATTCAG
TCACCTTATTCATGATTAAATGACTTATATCGACTAATCACAAAGATATCGGGACCTTATATTTTTT
ATTCGGAATATGAGCAGGCATGTTAGGGACATCTATAAGATGGATTATCCGCATTGAATTGAGTC
AACCAGGATCGTTTATTGGTGATGATCAAATTTATAATGTAATTGTAACTGCTCACGCTTTCGTAA
TAATCTTCTTCATAGTTATACCAATCATAATTGGAGGATTTGGAAATTGATTAGTACCTTTAATGAT
TGGAGCTCCCGATATAGCATTCCCACGGCTAAATAATATAAGATTTTGATTACTACCTCCAGCGCT
GATACTTCTTATAACAAGAAGAGTGTCCAGCGCTGGAGCCGGAACTGGATGAACAGTTTACCCCC
CTCTATCAGGGAATATCGCTCATATGGGGTATTCTGTAGACTTTGCAATTTTTAGATTACATTTGGC
AGGTGTAAGATCAATTTTAGGGGCAATCAATTTTATTACTACTATTTTAAATATACGGCCAGCCGG
AATAACCCTAGAACGTACTCCTTTATTTGTATGATCTGTAGGAATTACTGCATTACTACTATTGCTA
TCCCTTCCTGTTCTTGCAGGTGCTATTACGATACTTTTAACTGATCGTAATTTTAACACTTCATTCTT
TGACCCTGTAGGGGGTGGAGACCCCATTCTTTATCAACATCTATTTTGATTTTTTGGCCATCCCGAA
GTGTATATTTTAATTCTACCAGGATTCGGACTTATTTCACATATTATCAGAAAAGAAAGAGGTAAA
AACGAAACCTTCGGGCCGTTAGGAATAATTTATGCAATATTAGCTATTGGCATTTTAGGATTTATT
GTGTGAGCACATCATATATTTACAGTCGGAATAGACGTGGATACGCGAGCATACTTTACTTCTGCA
ACTATAATTATCGCAGTCCCAACTGGAATTACTTTTTTAGCTGATTAGCTACACTTCATGGTAGAA
AAATAACATCTTCCATTATTATGGAGCTCCCGAATTGTTCCCTGAAAA

Missing sequence here

CATCTGGCCGTATTTCTAAGAGATAGGTAGGGTATTAATGACTAGGATAGATATTATCAATAGTAG
ATGAAGCAAGCTCAAGCGCTCGTGGCGCACAAGGCTGTTAGTAGTGATGTGGTATCGCTTCTTCTT
TTTATCATTCACGATTGAAAGTCCCTTATCCCCTGTGACTATGTCCTACCGGAAAGTCACAAAGGG
ATGAGGTGGGCGGCGGATTCTTCTAGGGGGTCGAGCGGTTGCGAGAAGACCAGGGCATGTGAAAT
CAGTTTAACTTTTCTCTTCTTAATCGAGAGTCGCATATCATTGGTTGGGGATTGTAAGGATAGACA
GAATCCGGTGACGATAGTCATACCCCGATTCAGTCCGTATGTGCATCGGGTGATATTACTTCTTAA
CTTTAGCACGATATGGTATGCCTGGGTTAAGAATATGGGGCGGGACTACCACCCATATCATATAGT
AGATTATTGTCCCTGACCTTTAACTGGATCAATCGGGGCATTAACCATAACTTTTGGAACAACTGT
ATGATTACATACAGGAGAAACTTTTTTAACAAAATTAGGAATTATAATCACCCTATTAACTATATA
CCAATGATGACGAGATATTATCCGAGAAAGAACATTTCAAGGACATCATACCTTAAAAGTTGCAA
AAAGAATAAAAATAGGAATAATTCTATTTATTATTTCAGAAATTCTGTTTTTTGTATCATTCTTTTG
AGCTTTTCTCCATAGCAGAATTGCCCCATCAATTGATATTGGCAATGTGTGACCACCAGCCGGAAT
TAAAACATTCAATCCCATAGAAATTCCACTATTAAACACAATAATTTTATTAAGATCAGGAATTAG
AGTAACATGAGCTCACCATAGAATGATAGAAAAGAAAATTACACAAGCATCAAGGGGGCTACTTA

CAACCGTAATATTAGGAATATATTTTTCTATATTACAAGCATACGAGTATATACAAGCAGAATTCA
CTATCTCAGATTCAGTTTACGGCTCATGTTTCTTTATAATCACGGGATTTCATGGGATTCATGTAAC
TATTGGAACTATATTCTTAACAGTATGCTTTAGTACGTAATATAAAAACTCATTTCTCAAGAACCC
ATCACTTGGGGTTTGAAGCTGCTTCATGGTACTGACATTTCGTAGACGTAGTTGAATAATTGTATA
CCTAATCATTATAGACGAGAAGAATATCGCTACCACTGAAGACAGAGCGTGCACCTAGCACACCG
TTGACAGAAGCAAGATGATGTATGATCACTTCAGACGATATCATAACATCATCAGATGACTATAC
CCTTAGATCAGACAGGGAGTGCAGAAGAAAGCAGTCGATGTTCGATGGGTCACTAAAAATGTCTT
GTAGACACGCTGAC

Missing sequence here

GCGGTTTGCGGTTGTGAGTAGTTTCGGATGTATTATATTGCATGGTTTTATCCAATATGGACATCTA
CCGGTCTAATTAATCTGGATGAAATGTGATAAGTAATTAATGCGAAACCTGACCATAGGTAAATTT
CTTCAACTCAACAAACTATTCAATAGTAAGATTTCACCCAATAAATTTAAAGAAGGAGGACTGCA
TATATTATTACATCTTAACAAAAACCAGAACAAGCTCATAGAAGGCATAATTGAAATAAGCCCCT
TATTAATAAATAATCTTCGTCTATGAGACCGCTCATAAACAATATTAGCCAAGACAAACAACCCTG
AAGAACACAATCCATGCCCTAATATTATAATAAAAGATCCTATTAAACCTCAAATATTAAAAGTA
AATAAACCACAAACAACCATTCCTATATGAGCAACTGAAGAATAAGCGATCAAACATTTAATATC
AGTCTGAAAAAGACAAACAAACCCAGAAATTAATGATCCTATAACTCCAACACTTATCAAATAAA
AAGAAATGCTTGTATAACAAGCAGACATAAAAGATATAACACGAAATAAACCATAACCTCCCAAC
TTCAACAACACACCTGCCAAAATTATGGATCCAAACACCGGAGCTTCTACATGCGCCTTAGGTAAT
CAGAAATGGAAAAAAAACATAGGCATTTTAACTAAAAAAGCTAAAATTATAGACAAATATAAAT
AAATTCTACCTGAACCCTCTAAAAGCCACCAAACAAAACTATTACTATGATATCCTACAAAAAGA
AGCCCTAATAAAAGGGGAAAAGAAGCTTGGTCACATTGCTTTCCATATCATATTCAGTCTATTGCA
AACCCGAGAAAGAGAACGCATAGTCAAATGTTCACTATAAGGAGATTATCCTATTGTAGTACAGT
TA

Missing sequence here

CGCTGTTATCCCTAAAGTAACTTAAACTTATAATCAGAATAAACTGGATCATAAAAACATTAATTA
ATGAAAATAATAAAATAAAGTTATATAAATTTACCTATCGCCCCAATAAAATTTGATAATAAAAG
CAGAAAAGAAACACCTATACATTTATACTTCTAAAGTCAAATAAAGTTTTAAAGGGTCTTATCGTC
TTCTAATAACATCTCTGCTTTTTCACAGAAAAATAAATTTATTTTAAATAAGATAAAGAAAGTTAA
CCCTTCGTTTGACCATTCATGCCAGTTTCCAATTAAAAAACAAATTATTATGCTACCTTTGCACAGT
TAAAATACTGCGGCCATTTAAAACTCATTGGGCAGGTCTGACCTTAAATAAAAACAAAAAGACAT
GTTTTTGATAAACAGGCG

APPENDIX D
Amino acid sequences of *Cimex adjunctus*

ND2
SXKMLFFIMMIMSXXLVLSSNNWLSIWMGLEINLMSFIPLMFNPKNMMISESCMMYFLIQSISSMLLIFS
ILIKPMXXLELMNITQYTIMISMMMKLGLPPFHYWFVNVMKKMEWLSCLVLMTWQKIAPMMIMSYV
DFKIIIAIIAAIVANITAINQTDLRKIMSLSSINHMGWMLACMKESNAWIIYLMLYMLMSTPLLIFFNKKK
ILFINQMNKLNTTSDKLTLSMLLMSMAGLPPFTGFLPKWMAIQYMIMNENIFMAFLMTTMSLIAVFYY
LRMINTMMLLNSDKSKWTLSHNKSYGSYMMMCNLLLPISLFLID

COI
MFFIRNMSSHVSMWAGMLGTSMSWIIRIELSQPGSFIGDDQIYNVIVTAHAFVMIFFMVMPIMNGGFGK
WLVPLMIGAPDMAFPRLNNMSFWLLPPALMLLMTSSVSSAGAGTGWTVYPPLSGNIAHMGYSVDFAI
FSLHLAGVSSILGAINFITTILNMRPAGMTLERTPLFVWSVGITALLLLLSLPVLAGAITMLLTDRNFNTS
FFDPVGGGDPILYQHLFWFFGHPEVYILILPGFGLISHIISKESGKNETFGPLGMIYAMLAIGILGFIVWAH
HMFTVGMDVDTRAYFTSATMIIAVPTGI

COIII
PWPLTGSIGALTMTFGTTVWLHTGETFLTKLGIMITLLTMYQWWRDIIRESTFQGHHTLKVAKSMKMG
MILFIISEILFFVSFFWAFLHSSIAPSIDIGNVWPPAGIKTFNPMEIPLLNTMILLSSGISVTWAHHSMMEK
KITQASSGLLTTVMLGMYFSMLQAYEYMQAEFTISDSVYGSCFFMITGFHGIHVTIGTMFLTVC

ND4
ASFPLLLGLLFVGYHSNSFVWWLLEGSGSIYLYLSMILAFLVKMPMFFFHFWLPKAHVEAPVFGSMIL
AGVLLKLGGYGLFRVMSFMSACYTSISFYLMSVGVMGSLISGFVCLFQTDIKCLIAYSSVAHMGMVVC
GLFTFNIWGLMGSFIMMLGHGLCSSGLFVLANIVYERSHSRSLFINKGLISIMPSMSLFWFLLSCNNMCS
PPSLNLLGEILLLNSLLSWSNLPMVSFRIN

APPENDIX E

Nucleic acid sequence data of *Oeciacus vicarius* (5' to 3')

GACGCTCCCGGACGACTAAGACAACCGGGATCTTTTATTGGAGACGACCAAATTTATAATGTAAT
CGTAACTGCTCATGCCTTCGTGATGATCTTCTTCATAGTTATGCCAATCATAATTGGGGGATTTGGG
AATTGATTAGTTCCATTAATAATTGGAGCCCCCGACATAGCATTCCCTCGTCTTAATAATATAAGT
TTTTGATTGCTCCCTCCATCATTAACACTATTATTAATTAGTAGAATAGCTGATATGGGAGTGGGT
ACTGGGTGAACTGTATACCCCCCCCTTTCCGGAAACATTGCCCACATAGGGTATTCAGTTGACTTT
GCAATCTTTAGATTGCATTTAGCAGGAATTAGATCAATTTTAGGGGCAATTAATTTTATCACCACT
ATTTTAAATATACGTCCTGCAGGTATGACTCTAGAACGGACCCCGTTATTTGTTTGATCTGTAGGA
TTAATTAGTAGAATAGCTGATATGGGAGTGGGTACTGGGTGAACTGTATACCCCCCCCTTTCCGGA
AACATTGCCCACATAGGGTATTCAGTTGACTTTGCAATCTTTAGATTGCATTTAGCAGGAATTAGA
TCAATTTTAGGGGCAATTAATTTTATCACCACTATTTTAAATATACGTCCTGCAGGTATGACTCTAG
AACGGACCCCGTTATTTGTTTGATCTGTAGGAATTACAGCATTACTATTACTTTTATCCCTACCTGT
ACTAGCAGGTGCAATCACTATGTTATTAACCGATCGCAATTTAAATACTTCGTTTTTTGACCCTGTT
GGAGGGGGGGATCCTATTCTTTATCAACATTTATTCTGATTCTTCGGGCATCCTGAAGTATATATTT
TAATTCTTCCAGGATTCGGGCTAATTTCCCATATTATTAGTAAAGAAAGAGGAAAAAACGAAACC
TTTGGGCCTTTAGGTATAATTTATGCTATGTTAGCAATTGGAATTTTAGGATTCATCGTATGAGCTC
ACCATATATTCACTGTAGGAATAGACGTAGACACTCGAGCATACTTTACTTCTGCAACTATAATTA
TCGCAGTCCCTACTGGTATTAAGATTTTTAGATGGTTAGCTACACCTTCACGGTAGAAAAATATCA
TTTACACCATCTATTATATGAGCATTAGGGTTTGTTTTTCTATTTACCATTGGAAGGAATTAACAGG
AA

Missing sequence here

ATGGGAAAGAAGGGCTCGGCAACCTTTCACAAAGAAAAAGAATTAGGATACGGCAGGTTGGAGA
ATCAATACCCAGTCATATCCTACCGGGTATGTTACGTCGGAGGCTGTTTTTGTTTAGTAGGGGTCA
TGCTATTGGAGTATGACGTCGTGGATCATAGTTTGAACTATCCTATGCATATGGTACTTGAGAAGG
GTCACGGTCTATGGGAGCTGGTCCGCCCTGTCGGAGGGCGTGGAGCCTCGCTCCATTTGAACGGTT
AACAGTGGGCTCGATGGACTGTGTGTTTGTTGGAGTTCGCGAGAGCCTGGGTGGTGATGCCTTTTT
TGTGGGCAACCCCCCTCACATACCTTATCCACCCTGTATTGTGTTACAACAGGTGTTCCGGACGTA
GAGGGTTCCAAACCAGGTGGAAATCGTATTCAAGTCTGTCAATTCCACCCCCTCCCCGTGATATGA
AGCACCCGATCTTCGCGGTAATACCACCCCCCTCGTTGAGGAGGTTAGGGTCTACCACCCATTACA
TATAGTAGATTACAGACCCTGGCCTTTAACAGGATCAATTGGAACAATAACTCTTACTTTCGGATT
AGTAATATGATTCCATAAAAAAGAAAGAATATTAATAATATTAGGAATTATTATTATTATATTGAC
CATATTTCAATGATGGCGAGATGTTATTCGAGAAAGTACCTTCCAAGGTCATCATACAATAAAGGT
TACTAAGGGCATCAAAATAGGAATAATTCTATTTATTATCTCTGAGATTTTATTCTTCGTATCATTT
TTCTGAGCTTTCCTACACAGAAGAATCGCCCCAACAATTGAAATTGGCAACACCTGACCCCCAACC
GGTATTAAACCGTTTAATCCTATAGAAGTACCCTTATTAAATACCATAATTCTATTAACATCGGGA
ATTAGAATCACATGAGCACATCACAGAATCATGGAAAGAAAACACTCGCAAGCTAGTAAGGGCCT
ATTAACAACCGTATTATTGGGAATTTTTTTCTCAATACTACAAGGGTATGAATACTACCAAGCCGA
ATTTACTATTTCAGACTCAATTTATGGATCATGCTTTTTTATAATTACAGGATTTCACGGTTTACAT
GTAATCATTGGAACTATATTTTTAATAGTATGTTTAATGCGGAATATAAAAAAACCATTTTTCTAA
ATTTCACCACTTAGGATTTGAGGCTGCTTCATGATATTGACACTTTGTAGACGTTGTATGACTATTC
GTATATATCATCATTTACTGATGAGGAAGATAACTTATTCTTTTAGTATATCAGTATATCTAACTTC
CAATTAGAAGGACTTTAAATAGAAAGAATAATTCTTATGATAATAATCATTATAACAACCGCATTA
ACTATTTCAGCTATTATAATGATACTGTGCATTATAACAAGAAAAAAAAGAAATACAGACCGAGA
AAAAATATCTCCATTCGAATGTGGATTTGACCCAAAAAGATCACCTCGAATACCATTTTCAATCCA
ATTCTTCTTGATTGCCACTTTATTCCTGGTATTTGATGTGGAAATTACAATTATTATTCCTATAATTA
TTACATTCAACATAACCTCAATGAAAATTTGATTAATAACAACAACAACATTTATCATAACTTTGT
TAATTGGTTTATACTATGAGTGGTATAACAACATATTAGAATGATCCTCTTAGGGAATGTAGTTAA
AAATAACATTTAATTTGCAATTAAAAGGTACTAAGCATTTTAGTCTTTCTCAATATAAGTAATGAA
GTAAAACTTACATTTAGTTTCGACCTAAAAAAAGAGATTAATATCTCCTTACTTTTAATAGAAGCC

AAAAAGAGGCGATCTATTGTTAATAGATTTATTGATACAGTATATCCTATTAAAAGAGAATGATTG
TAACTAAAAGAAGCTGCTAACTATCTTTTAAAGCGGTTAAAATCCGTTTTTCTCTTATTTAAGTAGT
TTATAAAAAACACTTCATTTTCAATGAAAAAAAAATTAATAATTCTTAAATATTAAAAGAGAAAT
AATCTCATATAAGCAGCTTCAATGCTTCGCTTTAAATTAAGCTATTTTAATTAAAAAAAGAAAAAT
AATGATCTTAAAATAAAAGATACTAAAATCAATTTAATACCATTATTATAAAACCCAACAGTATTA
GCCGAAATCTTAACTAGAAAAAGAAATACAGACCGAGAAGCCAGATACTCCCCTCAACTACAATC
AATATTGGTATACCAAACCCTAGATATATATAACCCAAAACTTCCAAAAAAACGAGTTCTAAAAT
AAGGTATAAATAATATAGAGCCAAAAACCACACTAATAAAGTAAAAATTACAATTTAATGAACAT
ATTATATATGAAATAACAGCTCTAAAAATAATAATAAACAAAGGCAACAATTTTACAAAAATACT
TAAACAATAAAGATCAGGTACTAAAAAGATTACTCAACTCAAAACGCCCCCGCCTAAAATTGATA
TTACAACCATAATAAATATAGAAGAAACTATAAAACCATCCTCAACATAACTTTGACAGCAATAT
AAATTATCACTACCTAAAGTATAATAAATCAAACGTATTGAATATGATATTGTCAACCCCACAGAA
ACATAATACAAAAAAATAATAAAATAATTGTTATAATTACATCTTATAGCCTCAAGAACCATATCC
TTAGAATAAAATCCCGATAAAAAAGGAAACCCGCACAAAGATATATTAGAAATTATAAAACAAGT
ATATGTCCAAGGAAGACTCCTATAAACTCCTCCTATCAAACGCAAATCCTGTCTATTAGCTGCAAC
ATGAATAATCAACCCCGCACATAAAAACAATAAAGCCCTAAAAAAAGCATGTATTAACAAATGAA
AAAAACTAAGAATAGGATAACCCATAAATAAAATACCAACCATAAAACCTAACTGACTCAAGGTT
GATAGTGCAATAACCTTCTTAAGATCAAATTCAAAAATTGCTCTTAGTCCAGCTAAAAATATTGTT
AAGGTACCAACATAAACGAAAAATTCACAGTCATAATTCATGAACAAATAATTAAAACGAATCAA
TAAATAAACTCCAGCAGTAACAAGAAGTAGATGAATGGAACTAAAAGCAGAGACATTCAGTAAA
AAACAACAACAGCAA

Missing sequence here

ATTGCAGCAGGCAATCAAGAAGAAAAGGGGATTTGAGCTCTCCTGGTAAATCTTGCTAACATAAC
TATTAATAATACTACAATTATATTAAAATCCCAAATCCCTACAAAAAAAAGATAATGTCAACCTCC
AAAATTAAATATTCAAGCCAAGGCCATTAATAAAGCAGCATCACCAATACGATTAATCAAAATTG
TCAACATTCCTGCATTATAGGATTTATAATTTTGAAAATAAATAACTAAACCATATGAAACAAGAC
CAAGACCGTCTCAACCCAACAAAATACTGATTAAACTAGGACTAATAATTATAAACAGCATTCTC
AAGACAAATATTAAAACTAAATATAAAAATCGAATATTATTTATATCGCCTGACATATAAGAAAA
TCTATATAAAATAACCATAGAAGAAATAACCAAAACAAAAGACATAAATATCAAAGCCATTCAAT
CTAATAATACAACAAAAGATAAACTAACAGAATTTACAGAAAGAAGCTCCCAATCTATAAAAATA
ACATAGCTATAATTTACAAAACAAAGCCCCAAAACAAATAAAAAGCAGGCAGCAATAAATAAAT
AAAACGAACTTAATAAATATAGTATTACTTATAACACAAGCGTGTACCTGTAGTCCCACAAACTA
ACATTCTAAATTAAACTATAATAAGTAAAAGAAAACAAACAAAAAATCGATCCTAAAAATAAAA
ATATTTAAGGGGAGCCAATGCATAAAAACTAATAAATACTCACGAATTCTAGCATTATAACCAAA
AAAAACACCCTTATATGCAGGACCATGCTGAGTTATGCAATATAAATAGATTCTGTAACAACAAC
TCATAAAAGAATTAATTATTAAAAAAACTATAGTAAATGTTCTCCAGCTTATGATGCTATTTAATA
AAAACACTTCCCCTAATAAATTTAAAGAAGGAGGTCTAGATATATTATTAGCTCTAAATATAAATC
AAAATAAACTTATTGAAGGCATAATCGAAATATAACCTTTATTAATAAACAATCTCCGTCTATGAG
AGCGTTCATAAACAACATTTGCCAAAACGAATAAACCAGAAGAACATAAACCATGACCTAATATT
AATAAAAAAGAACCTATCAGACCCCACAAATTAAAAGTAAATAGACCACAAATCACTATACCCAT
ATGAGATACCGAAGAATATGCAATCAAACACTTGATATCAGTCTGAAATAAACAAATAAAAGCAG
CCAATACAGAGCCTAAAAGACCAATACTAACAAAAATAAAAGTGGCTTCAATATAATACTTATAC
ATAAAAAATATAACCCGAATAATGCCATACCCACCTAACTTTAATAGCACACCAGCTAGAATCAT
AGATCCAAAAACAGGAGCCTCAACATGAGCTTTAGGTAATCAAAAATGAAAGATAAATATAGGTA
TCCTAACTAAAAAGGCCAAAACCAAAGATAATTGTAAATAAAAACTATCAAACCCAACCAAAAAC
CACCAAATAAAACTACCTGATATAAATCCTAAAAAACAAATACCCAGTAAAAGGGGCAAAGAAG
CAAAAAGTATATAAAACAACAAATAAAACCCAGCATTTAAACGTTCAGGCTGGTACCCTCAACCA
AAAATAAGAAACAAGGTCGGAATTAGAGTTGATTCAAAAAAGACAAAAAATAAAAATATATTAA
CAGTTATAAAACAAAAGACCAAAAACAAAAGTAACAAAACACAAAGAAACAAAAAATCACTAAC
AGAACCAGAAGACTTTAATCTAAAACCTGAAAGAATTATTAAAACCACAATTCACAAAGACAAAT
AAACTATACACCAAGAAAAGGTATCTCCACCAAAAGAGAATGATAATACACTAAAATAAGGTATA

AACCAATTACTAAAAAAACTAACAAAGAATAAAACCAAAACAAAATTTATTACTTCTCAATAAAG
TCCCAGTAAAGCCACGGGGATCAAAAACAATAAAGAACAAATTATTATCATATTAACAAAGATAT
ACTCAATAAATTATCATTGCCGTGGCAACGAATTAATCTAACCAAAAGAGATAAACCCAAAACAC
CCTCACAAACAGTAAATGACAAATAAATTAATACAAAATAATGACCAGAACCTAAGTCAAAAAA
AAAGACAAAAAACATTAAAAATAAAAGCAAAACTATATATTCTAATCTAAGTAATATTAAAAGTA
AATGCTTATAAACCCTACAATAAACAAATAGACCTATTATAAATAAATAAGACAAAAAAACAAAC
ATGAGTTTTAATAGTTTAAGAAAAATAATGATCTGTGAATTAGGGGGGGGGGGGGTCCTGTAGTG
TGGAGGGCCGTCGCTGTGATAGGCACCGAATAAAAAACACACTTCTACGGGTTGCTATAGGGATA
TG

Missing sequence here

TTAATATTGAAAATAGCTTTCGAATAGATTAATTCATATCATACGTGACGTTAATAAATGGTTGAT
TATTACGAAGCATGCATGCTAACGGAGCCTCGTTTTTTTTCATTTGCATATACCTTCATGTAGGTCG
TGGAATTTATTATAATTCTTTTCAACTAACTCATACCTGAATAATTGGTGTTATACTTCTACTAATA
ACCATAGCTACAGCCTTCCTAGGATATGTTTTACCTTGAGGGCAAATATCTCTCTGGGGGGCGACA
GTCATTACAAACTTGCTTTCTGTTATTCCTTATATTGGTAATAATATTATTATATGATTATGAGGGG
GATTCTCAATTGAAAACGCCACATTAACACGATTCTTTACAATTCACTTTCTGTTACCATTTATAAT
TTTAATATTTGTAATTATTCACTTATTATATTTACACCAGACAGGATCAAACAACCCTATAGGGATT
AATAGAAATACTGATAAAATCCCATTCCACCCTTATTTTACTGTAAAAGATAGAATAGGTTGCATA
ATGATAATATTGCTATTTTTATTATTAAATATAATAGAACCTCAAATATTAGGAGACCCAGAAAAC
TTTATTCCCGCTAATCCCTTAGTTACTCCTATTCATATTCAACCAGAATGATATTTTCTATTTG

Missing sequence here

GAACAATTACCGCCCAAAATCTTTTCGACAAGCCCTTCGGCATGAGCGCACGGCTGGCTAAAGTT
GGGCTTAGCGTCTACTAAGAACATCTCTGCTTTTTCACGGAAAAATAAAATTAATTTATATGAGAT
AAAGAAAGCCAATCCCTCGTATGACCATTCATGCAAGTTTCCAATTAAAAAACAAATTATTATGCT
ACCTTTGCACAGTTAAAGTACTGCGGCCATTTAAAATTCATTGGGCAGGTCTGACTTTAAATAAAA
TCAAAAAGACATGTTTTTGTTAAACAGGCGAGGATTAGATTTGCCGAATTACTATATCTTAATTAT
AAAAATACTAATTCAATCATTATTACAATTTCAATACTAATTTATGAAATAATCTGAATATTTAAA
TAAATAACTATAAAATTACATTAAACTTAAAACAAACTATAACGAAATTAACGATGGAAGTTTCT
AATCCTACAAAATACAAGTAACATACAAGCATTATATAATACTAACAAGCTAATCCTTATTAGTCT
TTCAATTTTATATTACTAAGAATATAATTTTGATAAAATATAAAATAGACAAATTAAATTCGAATC
TCCAGAAACCAGATATATTGTGAACGAAAGGCATATAACCTCTATATGTTTATTCAAGATACTAAT
GTAACAGCAACCTTCATAAAAAAATATATCTCACAATTTCGGGATAAATTAATATATTAAATATAT
TTTGATTAACCCTGATACAAAAGGTACTTCATAAAAGAATTCTTTCCAAAAGTATATGTATTTTCC
CTCACAGTACTAATTCAATATTATTTATATTAAATCTATTCGATAAACCCTATACTAAGAATATAAT
TATTTTTTTTAAATTTAAAAGAAAAAGAATAACTATAAATAATATACTTTCAAATTAAGCTGATTC
AAACAGCTCACTTATCAATGTAAATGATAAAACTACTCAGTAGTTATAATTTGTTATAATCAAGAT
TTACTTTCCAGTAAATCTACTTTGTTACGACTTATCTTGCATCCAAGAGCGACGGGCGATGTGTGC
ATAATCTAGAGCTCAAATCATAATTAAAAGATATTAACTATTACTACCAAATCCTAATTCAATGAA
AATGTTTCATTACCAAATCCTAAATTTATAATAAAGTAACCCACCACTAGCTCTTTTAGTGCTGCTC
CTTGACCTGATATTTTCCCCATAAAGAAAGACGAAAATCAACCTTATAATACATTCAATAACAGAG
GTATACAAGCTAAAAAAAAGTAAGATGCCTCGTGGATAATAAATACCGCACCCACCCAGG

APPENDIX F
Amino acid sequences of *Oeciacus vicarius*

COI
SMSWIIRIELSQPGSFIGDDQIYNVIVTAHAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRLNNM
SFWLLPPSLTLLLISSMADMGVGTGWTVYPPLSGNIAHMGYSVDFAIFSLHLAGISSILGAINFITTILNM
RPAGMTLERTPLFVWSVGITALLLLLSLPVLAGAITMLLTDRNLNTSFFDPVGGGDPILYQHLFWFFGH
PEVYILILPGFGLISHIISKESGKNETFGPLGMIYAMLAIGILGFIVWAHHMFTVGMDVDTRAYFTSATMI
IAVPTGIKIFSWLAT

COIII
EVSVYHPLHMVDYSPWPLTGSIGTMTLTFGLVMWFHKKESMLMMLGIIIIMLTMFQWWRDVIRESTF
QGHHTMKVTKGIKMGMILFIISEILFFVSFFWAFLHSSIAPTIEIGNTWPPTGIKPFNPMEVPLLNTMILLT
SGISITWAHHSIMESKHSQASKGLLTTVLLGIFFSMLQGYEYYQAEFTISDSIYGSCFFMITGFHGLHVIIG
TMFLMVCLMRNMKKPFFISPLSIWGCFMMLTLCSRCMTIRMYHHLLMSK

ND3
MESMILMMMIIMTTALTISAIMMMLCIMTSKKSNTDREKMSPFECGFDPKSSPRMPFSIQFFLIATLFLV
FDVEITIIIPMIITFNMTSMKIWLMTTTTFIMTLLIGLYYEWYNNMLEWSS

ND5
MLYLLSSFYLFIAACFLFVLGLCFVNYSYVIFMDWELLSVNSVSLSFVVLLDWMALMFMSFVLVISSM
VILYSFSYMSGDMNNIRFLYLVLMFVLSMLFMIISPSLISILLGWDGLGLVSYGLVIYFQNYKSYNAGM
LTILINRIGDAALLMALAWMFNFGGWHYLFFVGIWDFNMIVVLLMVMLASFTSSAQIPFSSWLPAAFS
SIHLLLVTAGVYLLIRFNYLFMNYDCEFFVYVGTLTMFLAGLSAIFEFDLKKVIALSTLSQLGFMVGILF
MGYPILSFFHLLMHAFFSALLFLCAGLIIHVAANSQDLRLMGGVYSSLPWTYTCFMISNMSLCGFPFLS
GFYSKDMVLEAMSCNYNNYFIIFLYYVSVGLTMSYSMRLIYYTLGSDNLYCCQSYVEDGFMVSSMFI
MVVMSILGGGVLSWVIFLVPDLYCLSIFVKLLPLFIIIFSAVISYMMCSLNCNFYFISVVFGSMLFMPYFS
TRFFGSFGLYMSSVWYTNIDCSWGEYLASRSVFLFLVKISANTVGFYNNGIKLILVSFILSSLFFFFLK

ND4
MMMICSLLFLIPVALLGLYWEVMNFVLVLFFVSFFSNWFMPYFSVLSFSFGGDTFSWCMVYLSLWIVV
LMILSGFSLKSSGSVSDFLFLCVLLLLFLVFCFMTVNMFLFFVFFESTLIPTLFLIFGWGYQPERLNAGFY
LLFYMLFASLPLLLGICFLGFMSGSFIWWFLVGFDSFYLQLSLVLAFLVSMPMFIFHFWLPKAHVEAPV
FGSMILAGVLLKLGGYGIIRVMFFMYKYYIEATFIFVSIGLLGSVLAAFICLFQTDIKCLIAYSSVSHMGM
VICGLFTFNLWGLMGSFLLMLGHGLCSSGLFVLANVVYERSHSRSLFINKGYISIMPSMSLFWFMFSAN
NMSSPPSLNLLG
EVFLLNSIMSWSTFTMVFLMINSFMSCCYSIYLYCMTQHGPAYKGVFFGYNASIREYLLVFMHWLPLN
IFIFSIDFLFVFFYLL

ND4L
MFVFLSYLFMMGLFVYCSVYKHLLLMLLSLEYMVLLLFLMFFVFFFDLGSGHYFVLIYLSFTVCEGVL
GLSLLVSLIRCHGNDNLLSMSLLMW

CytB
MNGWLLRSMHANGASFFFICMYLHVGRGIYYNSFQLTHTWMIGVMLLLMTMATAFLGYVLPWGQM
SLWGATVITNLLSVIPYIGNNIIMWLWGGFSIENATLTRFFTIHFLLPFMILMFVIIHLLYLHQTGSNNPM
GINSN
TDKIPFHPYFTVKDSMGCMMMMLLFLLLNMMEPQMLGDPENFIPANPLVTPIHIQPEWYFLF